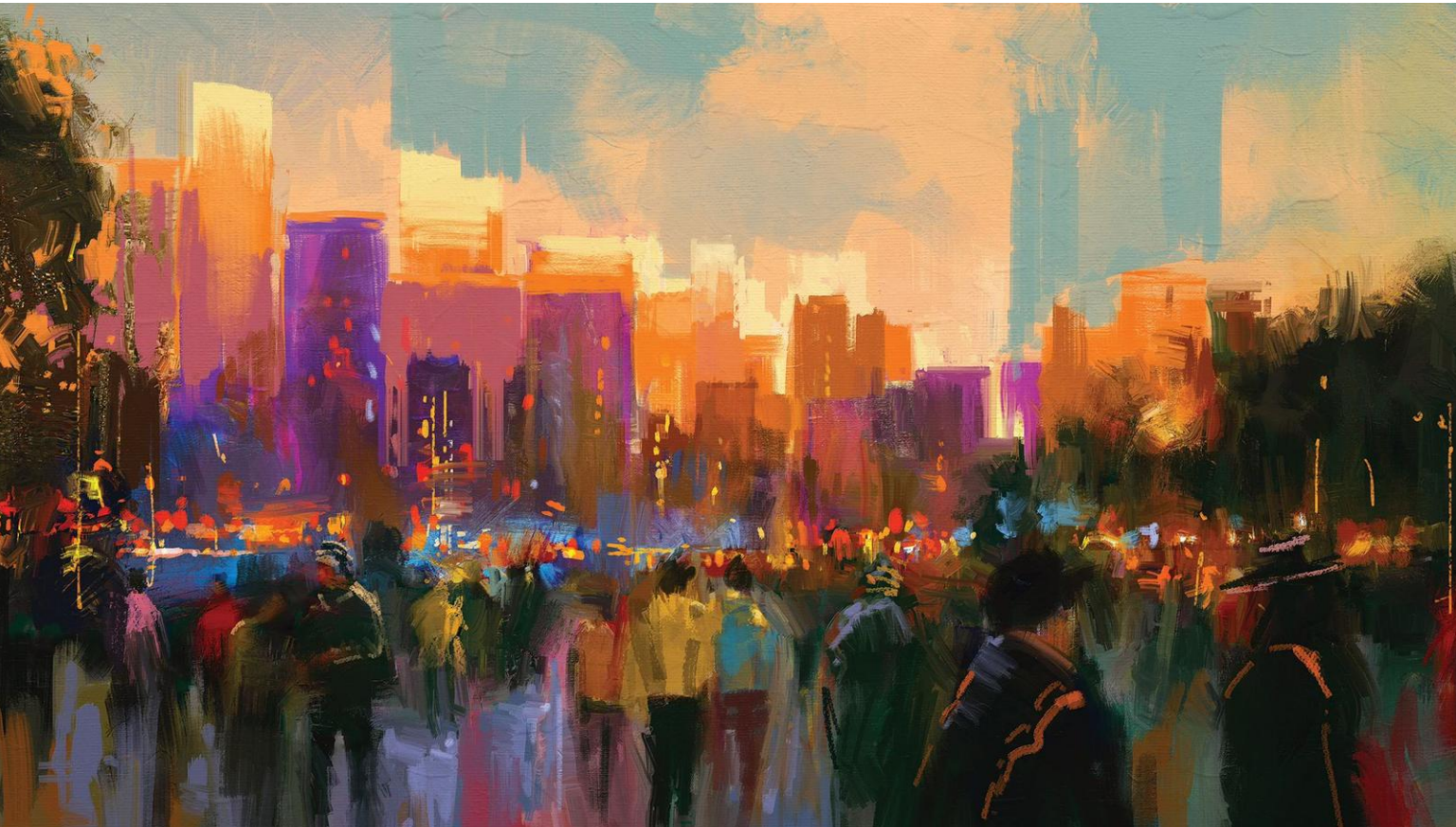


Épidémiologie appliquée

3^e édition

Une initiation à la lecture critique en sciences de la santé



Sous la direction de **André Simpson** • **Clément Beaucage** • **Yv Bonnier Viger**

CHENELIÈRE
ÉDUCATION



Épidémiologie appliquée

3^e édition

Une initiation à la lecture critique en sciences de la santé

Sous la direction de **André Simpson • Clément Beaucage • Yv Bonnier Viger**

Achetez
en ligne ou
en librairie

En tout temps,
simple et rapide!
www.cheneliere.ca

CHENELIÈRE
ÉDUCATION

Épidémiologie appliquée

Une initiation à la lecture critique en sciences de la santé, 3^e édition

Sous la direction de André Simpson, Clément Beaucage
et Yv Bonnier Viger

© 2017 TC Média Livres Inc.

© 2009 Les Éditions de la Chenelière inc.

© 1996 gaëtan morin éditeur ltée

Conception éditoriale: Dominique Hovington

Édition: Renée Théorêt

Coordination: Johanne Lessard

Révision linguistique: Anne-Marie Trudel

Correction d'épreuves: Dominique Poirier

Adaptation de la conception graphique originale: Micheline Roy

Conception de la couverture: Eykel design

Catalogage avant publication de Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada

Vedette principale au titre:

Épidémiologie appliquée: une initiation à la lecture critique de la littérature en sciences de la santé

3^e édition.

Comprend un index.

ISBN 978-2-7650-5334-7

1. Épidémiologie – Recherche – Méthodologie. 2. Épidémiologie – Méthodes statistiques. 3. Biométrie. i. Simpson, André, 1959- . ii. Bonnier Viger, Yv, 1950- . iii. Beaucage, Clément.

RA652.E64 2017

614.4072

C2017-940098-3

CHENELIÈRE
ÉDUCATION

5800, rue Saint-Denis, bureau 900

Montréal (Québec) H2S 3L5 Canada

Téléphone : 514 273-1066

Télécopieur : 514 276-0324 ou 1 800 814-0324

info@cheneliere.ca

TOUS DROITS RÉSERVÉS.

Toute reproduction du présent ouvrage, en totalité ou en partie, par tous les moyens présentement connus ou à être découverts, est interdite sans l'autorisation préalable de TC Média Livres Inc.

Toute utilisation non expressément autorisée constitue une contrefaçon pouvant donner lieu à une poursuite en justice contre l'individu ou l'établissement qui effectue la reproduction non autorisée.

ISBN 978-2-7650-5334-7

Dépôt légal : 1^{er} trimestre 2017

Bibliothèque et Archives nationales du Québec

Bibliothèque et Archives Canada

Imprimé au Canada

1 2 3 4 5 M 21 20 19 18 17

Gouvernement du Québec – Programme de crédit d'impôt pour l'édition de livres – Gestion SODEC.

Ce projet est financé en partie par le gouvernement du Canada

Canada

Sources iconographiques

Page couverture: Grandfailure/iStockphoto ;
p. 2, p. 16, p. 44, p. 72, p. 86, p. 106, p. 126, p. 152,
p. 172, p. 198, p. 218, p. 242, p. 258, p. 286, p. 320,
p. 344: HelgaLin/Shutterstock.com ;
p. 3, p. 17, p. 45, p. 73, p. 87, p. 107, p. 127, p. 153,
p. 173, p. 199, p. 219, p. 243, p. 259 ; p. 287,
p. 321, p. 345: Serge Gaboury.

L'éditeur tient à remercier chaleureusement Sylvie Cardin, du Département de médecine sociale et préventive de l'Université de Montréal, Mélissa Généreux, du Département des sciences de la santé communautaire de l'Université de Sherbrooke, Caroline Sirois, du Département des sciences infirmières de l'Université du Québec à Rimouski, René Verreault, du Département de médecine sociale et préventive de l'Université Laval, pour leur participation à l'évaluation de l'ouvrage.

Les cas présentés dans les mises en situation de cet ouvrage sont fictifs. Toute ressemblance avec des personnes existantes ou ayant déjà existé n'est que pure coïncidence.

Tous les sites Internet présentés sont étroitement liés au contenu abordé. Après la parution de l'ouvrage, il pourrait cependant arriver que l'adresse ou le contenu de certains de ces sites soient modifiés par leur propriétaire, ou encore par d'autres personnes. Pour cette raison, nous vous recommandons de vous assurer de la pertinence de ces sites avant de les suggérer aux étudiants.

Le matériel complémentaire mis en ligne dans notre site Web est réservé aux résidents du Canada, et ce, à des fins d'enseignement uniquement.

L'achat en ligne est réservé aux résidents du Canada.

Avant-propos

La genèse du présent ouvrage remonte à la fin des années 1980, époque où s'est imposée l'évidence qu'une meilleure gestion des ressources dans le monde de la santé devait constituer une priorité à la fois politique, économique et sociale. Parallèlement à ce mouvement de société, une autre obligation est née : faire le point sur les connaissances en distinguant celles qui sont essentielles et actuelles de celles qui sont secondaires ou désuètes.

Pour satisfaire à ces obligations, il apparaît clair que les étudiants universitaires de tous les horizons des sciences de la santé doivent désormais acquérir des connaissances de base en biostatistique, en épidémiologie et en analyse critique de la littérature scientifique. En premier lieu, ces connaissances leur seront nécessaires pour répondre aux exigences scolaires sanctionnées par les examens ; en second lieu – et c'est là sans doute le plus important –, elles se révéleront indispensables dans leur pratique.

S'appuyant sur 20 années d'utilisation par des étudiants majoritairement de premier cycle universitaire dans différentes disciplines des sciences de la santé, la troisième édition du manuel intègre toutes ces notions de façon vulgarisée. Son contenu est présenté dans un format papier et dans un format numérique, adapté notamment à la formation en ligne. De surcroît, cet ouvrage constitue un outil fort utile pour les professionnels de la santé déjà engagés dans leur pratique et qui sont souvent appelés à modifier leurs façons de faire pour tenir compte des progrès de la recherche et des changements sociaux. En fait, il est conçu de manière à permettre, à quiconque le souhaite, d'acquérir des connaissances dans le domaine par autoapprentissage.

Afin de répondre d'abord et avant tout aux besoins de l'utilisateur, l'ouvrage est écrit dans une langue simple, claire et précise. Autant que possible, les termes techniques ou scientifiques ont été évalués et remplacés par des termes plus courants. Un glossaire complet des termes utilisés vient soutenir le lecteur. Au surplus, pour en éprouver la forme et l'intelligibilité, les chapitres du manuel ont fait l'objet d'une évaluation par des étudiants en sciences de la santé. En définitive, on peut affirmer que les connaissances présentées sont adaptées, sur le plan de l'étendue et de la difficulté, aussi bien à une clientèle universitaire qu'à des cliniciens. L'une y trouvera un outil d'apprentissage sur mesure, et les autres, un ouvrage de référence facile à consulter et qui répondra aux interrogations survenant en contexte clinique.

L'objectif premier du manuel consiste donc à initier l'étudiant, ainsi que tout autre utilisateur, à la lecture critique de la littérature en sciences de la santé, une activité qui nécessite certaines connaissances en biostatistique et en épidémiologie. L'approche pédagogique retenue privilégie le raisonnement plutôt que la seule mémorisation. Ainsi, pour que le lecteur puisse transposer les notions expliquées dans un contexte pratique, la matière est présentée par l'intermédiaire de situations cliniques.

Le contenu de l'ouvrage

Le manuel comporte 16 chapitres: un premier chapitre présente un tour d'horizon de l'épidémiologie dans le contexte de différentes disciplines en sciences de la santé; cinq traitent de l'analyse critique de la littérature scientifique; neuf portent sur des notions en matière d'épidémiologie et de biostatistique; un traite de la prévention clinique. De plus, s'ajoute un nouveau chapitre disponible exclusivement sur le site Web du volume, qui traite de l'approche qualitative en épidémiologie. Ces blocs thématiques ne sont pas présentés de façon monolithique. Les différents chapitres sont répartis de façon que le lecteur soit rapidement amené à voir les implications cliniques des notions plus abstraites exposées, une formule qui permet un apprentissage «raisonné», intégré et graduel.

Une structure qui facilite l'apprentissage

Comme plusieurs auteurs représentant diverses facettes du domaine de la santé (praticiens, professeurs provenant de différentes universités, chercheurs, gestionnaires de la santé) signent l'ensemble des textes que comporte le manuel, il importait tout particulièrement de lui donner cohérence et homogénéité, lesquelles se situent sur deux plans: textuel et visuel. Ainsi, chacun des chapitres est rédigé en fonction d'un plan déterminé, d'une structure uniforme pour l'ensemble du volume.

Dans un premier temps, le lecteur trouve au début de chaque chapitre un tableau qui présente les habiletés à acquérir et précise les objectifs de même que le contenu détaillé qui y est associé.

Tous les chapitres s'ouvrent sur un cas constituant une mise en situation concrète d'un problème de santé actuel, à partir duquel les notions sont présentées; le lecteur peut donc saisir d'emblée le lien entre la matière qui fait l'objet du chapitre et la pratique clinique concrète.

L'enchaînement se fait avec une section intitulée «Pertinence clinique», un texte bref qui lie de manière plus explicite la mise en situation et les notions traitées dans le chapitre. Les problèmes ou les questionnements qui pourraient, voire devraient, se poser au clinicien sont ici soulevés en relation avec le cas présenté.

Vient ensuite le corps du chapitre, la «substantifique moelle». Que le chapitre porte sur des sujets d'ordre épidémiologique ou biostatistique, ou qu'il y soit question de lecture critique, l'exposé de la matière respecte une progression en ce qui concerne les niveaux de difficulté. Dans la mesure du possible, les concepts abstraits sont illustrés à l'aide d'exemples concrets, fictifs ou réels, qui facilitent une transposition dans la pratique. De plus, une attention particulière est accordée aux considérations éthiques.

Enfin, chaque chapitre se termine par une section intitulée «En résumé», où sont récapitulés les points essentiels de la matière présentée; elle est suivie de quelques questions d'exercices auxquelles les lecteurs sont invités à répondre avant de poursuivre avec les exercices en ligne accessibles sur le site Web du volume. Les «Notes et références» regroupées à la fin du volume fournissent pour chaque chapitre des

renseignements utiles ou des références aux articles ou à des ouvrages scientifiques récents et des hyperliens d'intérêt au lecteur désireux d'approfondir la matière.

Pour renforcer l'unité formelle de l'ensemble, une approche narrative a par ailleurs été privilégiée; ainsi, un scénario de base a été retenu. D'un chapitre à l'autre, le lecteur est invité à partager le quotidien de Lisa, une omnipraticienne à la polyclinique Milo, et celui des principaux personnages qui gravitent autour d'elle: Danielle, une collègue infirmière, Marie-Louise, une pharmacienne à la polyclinique, Jean, un ami épidémiologiste, et Stéphanie, une nutritionniste. Ces personnages, ainsi que les situations dans lesquelles ils sont appelés à évoluer, illustrent les préoccupations et les champs d'intérêt de divers intervenants du domaine de la santé.

Par ailleurs, afin de donner vie et couleur à cet univers, des bandes dessinées ont été conçues par le bédéiste Serge Gaboury. Ces bédés, qui viennent coiffer chacun des chapitres, s'inspirent directement du cas présenté en début de chapitre.

Des outils de travail additionnels sont mis à la disposition du lecteur dans l'ouvrage, soit une liste des symboles et des abréviations utilisés, le corrigé des exercices à la fin des chapitres, un glossaire ainsi qu'un index. Les termes en gras dans le texte sont tous définis dans le glossaire à la fin de l'ouvrage.

Nouveau

En outre, sur le site Web de l'ouvrage, le matériel complémentaire suivant est offert: l'ouvrage numérique, des hyperliens d'intérêt, le glossaire et un nouveau chapitre sur les méthodes qualitatives. Qui plus est, afin de permettre au lecteur d'évaluer la mesure dans laquelle il a assimilé les connaissances présentées, on y trouve une série d'exercices accompagnés de la mention de leur objectif ainsi que de leur corrigé pour le professeur. La version numérique du volume en fait aussi un outil bien adapté aux technologies actuelles et à la formation en ligne.

Conclusion

En voulant intégrer, tant sur le plan formel que sur le plan conceptuel, ces fondements des sciences de la santé que constituent l'épidémiologie, la biostatistique et l'analyse critique de la littérature, une constante préoccupation de simplification et de vulgarisation a dû être maintenue, ce qui ne signifie nullement un manque de rigueur et une concession aux impératifs scientifiques. Ce souci traduit en fait une volonté marquée de fournir un contenu qui soit directement assimilable dans un contexte d'autoapprentissage. La lecture de ce manuel devrait amener le lecteur à appréhender des notions souvent complexes en usant de sa capacité de raisonnement, ce qui, du coup, devrait lui permettre de mesurer l'importance de fonder sa pratique clinique sur de solides bases scientifiques.

Équipe de rédaction

Le travail de mise à jour pour produire cette 3^e édition a été rendu possible grâce à la collaboration de tous les membres de l'équipe. La contribution de chacun dans les deux éditions précédentes est mentionnée lorsqu'elle diffère de celle de cette présente édition.

Les directeurs

André Simpson est titulaire d'une maîtrise en épidémiologie et d'une maîtrise en sociologie de l'Université Laval. Il cumule 23 années d'enseignement en épidémiologie à l'Université Laval, 5 années à l'Université de Montréal et 3 années en santé et développement social à l'Université du Québec à Rimouski. Parallèlement, il a œuvré pendant 30 ans comme gestionnaire et épidémiologiste principalement dans un centre de dépistage et à l'Institut national de santé publique du Québec, dans les domaines du dépistage, de la surveillance de l'état de santé de la population, de la gestion de l'information et de la formation. En plus d'avoir collaboré à la rédaction du chapitre 9 de la 1^{re} édition, il s'est joint à Clément Beaucage et à Yv Bonnier Viger pour diriger les 2^e et 3^e éditions, dont il a rédigé le chapitre 1 ajouté à la 2^e édition. Il a également participé à la mise à jour des chapitres 1, 2, 5, 6, 8, 9 et 16 dans la 2^e édition auxquels s'ajoute le chapitre 4 dans la 3^e édition.

Clément Beaucage est médecin spécialiste en médecine préventive et titulaire d'un baccalauréat en pédagogie de l'Université Laval. Il pratique en santé communautaire au Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale et il est aussi professeur de clinique à l'Université Laval. Ses champs d'intérêt sont la santé mentale en milieu de travail et la prévention des risques à la santé chez la travailleuse enceinte. Soulignons également sa présence à la direction de l'ouvrage dans les trois éditions de celui-ci et sa collaboration à la rédaction des actuels chapitres 7, 8, 10, 13, 14 et 16 dans la 1^{re} édition de l'ouvrage et à la mise à jour des chapitres 1 et 16 des deux dernières éditions.

Yv Bonnier Viger, médecin spécialiste en santé publique et médecine préventive, est professeur au Département de médecine sociale et préventive de la Faculté de médecine de l'Université Laval et directeur régional de santé publique de la Gaspésie Îles-de-la-Madeleine. Titulaire de baccalauréats en sciences de la santé et en mathématiques et de maîtrises en épidémiologie et en gestion, il poursuit des études en pédagogie médicale. Soulignons également sa présence à la direction de l'ouvrage dans les trois éditions de celui-ci et sa collaboration à la rédaction des actuels chapitres 2, 3, 5, 7, 8, 13 et 14 dans la 1^{re} édition de l'ouvrage.

Les collaborateurs

Claude Julie Bourque est sociologue et titulaire d'un doctorat en éducation. Elle est coordonnatrice scientifique de l'Unité de recherche en éthique clinique et partenariat famille au Centre hospitalier universitaire Sainte-Justine. Elle se spécialise dans le développement de designs méthodologiques adaptés aux enquêtes complexes et aux terrains fragiles comprenant des devis qualitatifs et mixtes ainsi que dans l'enseignement et la formation à la recherche scientifique. Elle s'est jointe à la 3^e édition de l'ouvrage et en a rédigé le chapitre 17, en ligne.

Agathe Croteau est médecin spécialiste en médecine du travail et titulaire d'un doctorat en épidémiologie de l'Université Laval. Elle pratique à l'Institut national de santé publique du Québec. Ses principaux champs d'intérêt sont les effets des conditions de travail des femmes enceintes sur leur grossesse, les synthèses systématiques ainsi que le domaine du transfert et de l'échange des connaissances. Elle a collaboré à la mise à jour du chapitre 15 dans les deux dernières éditions.

Pierre Deshaies est médecin spécialiste en santé publique et médecine préventive et titulaire d'une maîtrise en épidémiologie. Après avoir pratiqué en médecine familiale pendant 10 ans, il œuvre à présent en santé publique à la Direction régionale de santé publique de Chaudière-Appalaches et à l'Institut national de santé publique du Québec. Chef du Département clinique de santé publique du Centre intégré de santé et de services sociaux de Chaudière-Appalaches, ses champs d'intérêt sont la santé et la sécurité au travail ainsi que la surveillance populationnelle. Par ailleurs, il est professeur de clinique à l'Université Laval et il agit aussi à titre de consultant dans des projets de coopération internationale. Il a participé à la rédaction des chapitres 4 et 9 dans la 1^{re} édition et à la mise à jour de ces deux mêmes chapitres dans les deux éditions suivantes.

Suzanne Gingras est titulaire d'une maîtrise en mathématiques, et elle a travaillé comme statisticienne dans le domaine de la recherche en santé publique pendant plus de 20 ans. À titre de chercheuse à l'Institut national de santé publique du Québec, ses champs d'intérêt étaient l'environnement et la surveillance de l'état de santé de la population. Elle a participé à la mise à jour des chapitres 10, 11 et 12 dans les deux dernières éditions.

Jean-Pierre Grégoire est professeur titulaire à la Faculté de pharmacie de l'Université Laval. Pharmaco-épidémiologiste, il est chercheur dans l'Axe Santé des populations et pratiques optimales en santé du Centre de recherche du Centre hospitalier universitaire de Québec – Université Laval. Ses activités de recherche visent à optimiser l'usage des traitements médicamenteux des maladies chroniques. Il a participé à la rédaction du chapitre 12 dans la 1^{re} édition et à sa mise à jour dans les deux éditions suivantes.

Danielle Laurin est titulaire d'un doctorat en épidémiologie de l'Université Laval. Elle a effectué une formation postdoctorale au National Institute on Aging des National Institutes of Health, à Bethesda, dans le Maryland. Elle est professeure à la Faculté de pharmacie de l'Université Laval et travaille à titre de chercheuse au Centre d'excellence sur le vieillissement de Québec. Elle a collaboré à la mise à jour des chapitres 2, 6, 13 et 15 dans les deux dernières éditions de l'ouvrage.

Antoine Lutumba Ntetu est titulaire d'une maîtrise en sciences infirmières, d'un diplôme en santé communautaire et d'un doctorat en andragogie de l'Université de Montréal; il est aussi diplômé en droit de l'Université de Sherbrooke. Il a été professeur et directeur du Département des sciences humaines à l'Université du Québec à Chicoutimi et il œuvre maintenant comme coordonnateur d'activités à l'Hôpital Sainte-Anne du Centre intégré universitaire de santé et de services sociaux de l'Ouest-de-l'Île-de-Montréal. Ses principaux champs d'intérêt pour la recherche

sont la santé des groupes vulnérables, la dimension culturelle de la santé, les maladies chroniques en soins de première ligne, les stratégies de formation en milieu de travail, la prévention et le règlement des différends. Il a collaboré à la mise à jour des chapitres 3 et 7 dans les deux dernières éditions.

Thomas Matukala Nkosi est titulaire d'un doctorat en santé publique-épidémiologie de l'Université de Montréal et a effectué un stage postdoctoral conjointement en médecine familiale et en épidémiologie clinique à l'Université McGill. Antérieurement chargé de cours dans plusieurs universités au Québec et en Ontario, il est maintenant professeur associé à l'Université Laurentienne (Sudbury, Ontario) et à l'Institut Supérieur des Techniques Médicales de Kinshasa en République démocratique du Congo. Depuis 2016, il est directeur général de l'Institut Supérieur des Techniques Appliquées de Kasangulu au Congo central. Il a participé à la mise à jour des chapitres 2, 5, 7 et 9 dans les deux dernières éditions.

Marie-Hélène Mayrand est gynécologue à l'Hôpital Saint-Luc du Centre hospitalier universitaire de l'Université de Montréal. Elle est également épidémiologiste et s'intéresse, tant par ses activités cliniques que de recherche, à la prévention des cancers gynécologiques. Elle a participé à la mise à jour des chapitres 7 et 14 dans les deux dernières éditions.

Jocelyne Moisan est professeure titulaire à la Faculté de pharmacie de l'Université Laval et chercheuse dans l'Axe Santé des populations et pratiques optimales en santé du Centre de recherche du Centre hospitalier universitaire de Québec – Université Laval. Elle enseigne la pharmaco-épidémiologie. Ses projets de recherche portent sur la qualité de l'usage des médicaments dans la population et sur les interventions visant l'amélioration de cet usage. Elle a participé à la rédaction du chapitre 12 dans la 1^{re} édition et à sa mise à jour dans les deux éditions suivantes.

Louise Moreault est médecin spécialiste en santé publique et médecine préventive et titulaire d'une maîtrise en épidémiologie de l'Université Laval. Elle est également professeure de clinique au Département de médecine sociale et préventive de la Faculté de médecine de l'Université Laval. Elle y enseigne notamment les notions de dépistage aux externes en médecine. Elle pratique actuellement en santé publique à la Direction de santé publique de la Capitale-Nationale. Ses principaux champs d'intérêt sont le dépistage du cancer du sein et les pratiques cliniques préventives. Elle a participé à la mise à jour du chapitre 16 dans les deux dernières éditions.

Pamphile Nkogh Mengue est microbiologiste médical et pharmaco-épidémiologiste, diplômé de l'Université Laval. Ses activités de recherche portent sur l'usage optimal des médicaments chez les aînés et sur l'épidémiologie des grands syndromes gériatriques. Il a participé à la mise à jour des chapitres 8 et 11 dans la deuxième édition et du chapitre 8 dans la dernière édition.

Helen Trottier est épidémiologiste, chercheuse au Centre de recherche du Centre hospitalier universitaire Sainte-Justine et professeure agrégée au Département de médecine sociale et préventive de l'Université de Montréal. Elle a participé à la mise à jour des chapitres 3, 7 et 14 dans les deux dernières éditions.

Remerciements

Nous tenons d'abord à remercier les rédacteurs et les collaborateurs des deux éditions précédentes qui ne font plus partie de l'équipe du présent ouvrage. C'est en grande partie grâce à leur travail de pionnier que cette troisième édition est possible.

Michèle Aubin, médecin et professeure au Département de médecine familiale de l'Université Laval (chapitre 16, 1^{re} édition); **Nicole Audet**, médecin et professeure adjointe de clinique au Département de médecine familiale de l'Université Laval (chapitres 4 et 15, 1^{re} édition); **Renée Bourbonnais**, professeure au Département d'ergothérapie de l'Université Laval (chapitre 12, 1^{re} édition); **André Dontigny**, médecin-conseil à l'Unité de santé publique du Centre hospitalier régional de Rimouski (chapitre 6, 1^{re} édition); **Chantal Guimont**, médecin et professeure au Département de médecine familiale et de médecine d'urgence de l'Université Laval (chapitres 10 et 11, 1^{re} édition, mise à jour des chapitres 11 et 13, 2^e édition); **Denis Laliberté**, médecin et professeur de clinique au Département de médecine sociale et préventive de l'Université Laval (chapitre 16, 1^{re} édition); **Louis-Gabriel Latulippe**, médecin et professeur au Département de médecine familiale de l'Université Laval (chapitres 13 et 14, 1^{re} édition); **Yvan Leduc**, médecin et professeur de clinique au Département de médecine familiale de l'Université Laval (chapitres 7 et 8, 1^{re} édition); **Pierre Mercier**, médecin et professeur de clinique à la Faculté de médecine de l'Université Laval (chapitres 10 et 11, 1^{re} édition).

Merci également aux étudiants en sciences de la santé de l'Université Laval, qui ont apporté leurs critiques et leurs suggestions à la première édition. Nous ne pouvons passer sous silence la contribution du linguiste Réal D'Amours à l'édition originale. Ses interrogations et ses suggestions ont teinté durablement la qualité de la langue de cet ouvrage. Merci également à nos collègues ainsi qu'aux auteurs qui ont pris part à la mise à jour de la troisième édition de ce volume. Soulignons par ailleurs l'apport de madame Denise Vigneault, du Réseau de valorisation de l'enseignement de l'Université Laval, dont les suggestions et les remarques, à la deuxième édition, ont contribué à enrichir la présentation et la révision des objectifs de chacun des chapitres.

Nous levons aussi notre chapeau à monsieur Serge Gaboury, bédéiste, créateur de la bande dessinée illustrant avec humour les thèmes clés de chacun des chapitres.

De même, nous ne pouvons passer sous silence le travail des professionnelles de la maison d'édition Chenelière Éducation à cette nouvelle mouture, soit mesdames Dominique Hovington, Renée Théorêt, Johanne Lessard, Dominique Poirier et Anne-Marie Trudel qui, par leur disponibilité et leur expertise, ont contribué à améliorer ce volume.

De plus, nous remercions du fond du cœur les membres de nos familles qui, avec grande générosité, nous ont encouragés tout au long de cette aventure, sans oublier celles de nos collaborateurs et collaboratrices.

André Simpson, Clément Beaucage, Yv Bonnier Viger

Liste sélective des abréviations et symboles utilisés

CV	Coefficient de variation	R	Coefficient de corrélation (de relation linéaire)
D	Durée moyenne	R²	Coefficient de détermination multiple
FE₁	Fraction étiologique chez les exposés	RA	Risque attribuable (différence de risque)
FE_t	Fraction étiologique totale	RC	Rapport de cotes
FP₁	Fraction prévenue chez les exposés	RI	Rapport de taux d'incidence
FP_t	Fraction prévenue totale	RR	Risque relatif
H₀	Hypothèse nulle	RRP	Risque relatif à la population
H₁	Hypothèse alternative	RV	Rapport de vraisemblance
I	Taux d'incidence	SE	Sensibilité d'un examen diagnostique
IC	Incidence cumulée	SP	Spécificité d'un examen diagnostique
ic	Intervalle de confiance	TA	Taux d'attaque
L	Létalité	VPN	Valeur prédictive négative
P	Prévalence	VPP	Valeur prédictive positive

$\sum_{i=1}^n x_i$ Somme des éléments x . Chaque élément x est identifié par un indice i . Ici, i varie de 1 à n .
Il y a donc n éléments à additionner.

Exemple:

$\sum_{i=1}^3 x_i$ est la somme de 3 éléments.

Si $x_1 = 12$, $x_2 = 20$ et $x_3 = 18$, alors $\sum_{i=1}^3 x_i = 50$.

$\prod_{i=1}^n x_i$ Produit des facteurs x . Chaque facteur x est identifié par un indice i . Ici, i varie de 1 à n .
Il y a donc n facteurs à multiplier.

Exemple:

$\prod_{i=1}^3 x_i$ est le produit de 3 facteurs.

Si $x_1 = 2$, $x_2 = 20$ et $x_3 = 10$, alors $\prod_{i=1}^3 x_i = 400$.

Table des matières

CHAPITRE 1

Un tour d'horizon de l'épidémiologie appliquée 2

1.1 L'apport de l'épidémiologie 3

1.1.1 L'épidémiologie selon une perspective populationnelle 4

La prévention de la maladie et les décès.. 5

L'intervention précoce et l'aide aux personnes atteintes 5

1.1.2 L'épidémiologie selon une perspective clinique 5

1.1.3 L'épidémiologie selon une perspective interdisciplinaire 6

La recherche à la polyclinique Milo 8

1.2 Les principes éthiques de la recherche et de l'intervention en santé 10

1.3 La lecture critique de la littérature en sciences de la santé 11

1.3.1 Les habiletés requises pour la lecture critique 11

1.3.2 La structure générale d'une publication scientifique 12

1.3.3 Un cadre de référence pour la lecture critique 13

En résumé 15

Exercices 15

CHAPITRE 2

L'approche statistique de la réalité 16

2.1 La notion de variable 17

2.1.1 Les types de variables 18

Les variables de personnes 18

Les variables de lieux 18

Les variables de temps 18

2.1.2 La classification des variables 18

Les variables quantitatives 19

Les variables qualitatives 19

2.2 La classification des observations 20

2.2.1 L'échelle de classification 20

2.2.2 Les types d'échelles de classification 21

L'échelle nominale 21

L'échelle ordinale 21

L'échelle par intervalle 22

2.2.3 Le passage d'une échelle à une autre 22

2.3 Les distributions de fréquences 23

2.4 Les représentations graphiques des distributions de fréquences 24

2.4.1 L'échelle nominale ou ordinale 24

Le diagramme en secteurs proportionnels 24

Le diagramme en barres proportionnelles 24

Le diagramme en barres 25

2.4.2 L'échelle par intervalle 25

L'histogramme 25

Le polygone de fréquences 26

Le graphe en lignes 26

Le graphe de percentiles 27

2.4.3 Les autres représentations graphiques 29

2.4.4 Les règles générales de représentation graphique 30

2.5 Les mesures descriptives générales d'un ensemble de données 30

2.5.1 Les mesures de tendance centrale 30

La moyenne 30

La moyenne géométrique 32

La médiane 33

Le mode 35

La relation entre les échelles et les mesures de tendance centrale 35

2.5.2 Les mesures de dispersion 36

L'étendue 36

Les quantiles 37

La variance 37

L'écart type 38

Le coefficient de variation 39

2.6 L'échantillonnage 39

2.6.1 Le choix d'un échantillon 40

2.6.2 Les méthodes d'échantillonnage 41

En résumé	42
Exercices	43

CHAPITRE 3

Les mesures et les modèles d'études en épidémiologie	44
3.1 Les mesures de fréquence	45
3.1.1 La proportion	46
3.1.2 Le taux	46
3.1.3 Le ratio	48
3.1.4 L'indice	48
3.2 Les mesures en épidémiologie	49
3.2.1 La prévalence	49
3.2.2 Le taux d'incidence	50
3.2.3 L'incidence cumulée	50
3.2.4 La cote	51
3.2.5 Une synthèse sur les mesures en épidémiologie	52
3.2.6 La population statique ou dynamique, ouverte ou fermée	52
3.3 Des mesures de mortalité et de morbidité	54
3.3.1 La mortalité	54
La probabilité de décès	55
La létalité	56
Un exemple de calcul de la létalité	56
3.3.2 La morbidité	56
Un exemple du calcul du taux d'attaque ...	56
3.3.3 L'espérance de vie	57
3.4 L'ajustement des mesures	59
3.5 Les modèles (ou types) d'études en épidémiologie	61
3.5.1 L'étude transversale	61
3.5.2 L'étude longitudinale	62
3.5.3 L'étude cas-témoins	63
3.6 Une classification des études épidémiologiques	64
3.6.1 Les études à visée étiologique	64
Un exemple d'étude transversale	64
Les études longitudinales	64
Les avantages et les désavantages des études de cohorte(s) et cas-témoins	66

3.6.2 Les études descriptives	67
L'étude descriptive de population transversale	67
L'étude descriptive de population longitudinale	67
Un exemple d'étude d'un cas ou d'une série de cas	68
L'étude écologique	68
L'étude de tendance	69
3.6.3 Une classification des types d'études selon leur finalité	69
En résumé	70
Exercices	71

CHAPITRE 4

La lecture critique des études descriptives : études d'un cas et d'une série de cas	72
4.1 L'étude d'un cas	74
4.1.1 Les caractéristiques de l'étude d'un cas	74
4.1.2 Les forces, les faiblesses et les limites de l'étude d'un cas	75
4.1.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'un cas (les exigences de publication) ...	75
La description du cas et le motif de consultation	76
L'anamnèse et l'examen clinique	76
La dimension chronologique	76
Le raisonnement et la prise en charge diagnostiques et thérapeutiques	76
Le suivi et l'issue des soins	76
La discussion	77
4.2 L'étude d'une série de cas	78
4.2.1 Les caractéristiques de l'étude d'une série de cas	78
4.2.2 Les forces, les faiblesses et les limites de l'étude d'une série de cas	80
4.2.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'une série de cas (les exigences de publication)	80
4.3 Une grille d'analyse des études d'un cas et d'une série de cas	81

4.4	Les enjeux éthiques pour les études d'un cas ou d'une série de cas	84
	En résumé	84
	Exercices	85

CHAPITRE 5

	Les mesures d'association et la causalité	86
5.1	Les mesures d'association	88
5.1.1	Le risque attribuable	88
5.1.2	Le risque relatif	89
	Le rapport de proportions (d'incidences cumulées ou de prévalences)	90
	Le rapport de taux (d'incidence ou de mortalité)	90
5.1.3	Le rapport de cotes	92
5.1.4	Le risque relatif à la population	94
	Les mesures d'association et les types d'études	95
5.1.5	Le coefficient de corrélation linéaire	95
	La signification du coefficient de corrélation	96
5.2	La causalité	98
5.2.1	L'association, la temporalité et la conséquence	99
5.2.2	Les critères de la causalité	99
	La temporalité	101
	La force de l'association	101
	La vraisemblance et la plausibilité biologique	101
	L'analogie	101
	La spécificité	101
	La relation dose-réponse	101
5.3	La mesure de l'impact d'un facteur d'exposition sur la fréquence des maladies et des décès	102
5.3.1	La fraction étiologique chez les exposés	102
5.3.2	La fraction prévenue chez les exposés	103
5.3.3	La fraction étiologique et la fraction prévenue totales	103
5.3.4	Le nombre de personnes à traiter	104
	En résumé	104
	Exercices	105

CHAPITRE 6

	L'utilité clinique des notions de précision et de validité	106
6.1	La justesse d'une mesure	107
6.1.1	La validité	107
6.1.2	La précision	109
6.1.3	Les relations entre la précision et la validité	110
6.2	Les types de biais	111
6.2.1	Le biais de sélection	112
	Le biais d'échantillonnage ou de survie sélective	112
	Le biais d'admission ou de Berkson	113
	Le biais de migration ou des perdus de vue	113
	Le biais de volontariat	114
	Le biais de « bonne santé des travailleurs »	115
	Le biais de publication	115
	La réduction des effets du biais de sélection	116
6.2.2	Le biais d'information	117
	L'erreur de classification non différentielle	118
	L'erreur de classification différentielle	118
	Le biais de rappel	118
	La réduction des effets du biais d'information	119
6.2.3	Le biais de confusion	120
	Le contrôle du biais de confusion	123
	En résumé	125
	Exercices	125

CHAPITRE 7

	Les examens diagnostiques	126
7.1	La classification des résultats d'un examen diagnostique	128
7.1.1	La comparaison d'un examen diagnostique à un examen de référence	128
7.1.2	La classification des résultats	129
7.2	La validité intrinsèque: sensibilité et spécificité	130
7.2.1	Le calcul de la sensibilité et de la spécificité	131

7.2.2	Le choix d'un examen sensible ou spécifique	132
	Le choix d'un examen sensible	132
	Le choix d'un examen spécifique	132
7.2.3	L'influence du seuil de positivité	132
	Le choix du seuil de positivité : l'exemple du dépistage du cancer du col de l'utérus	134
	La courbe récepteur opérateur	136
7.3	La validité extrinsèque: les valeurs prédictives positives et négatives	137
7.3.1	Les valeurs prédictives positives et négatives.....	138
7.3.2	Le calcul des valeurs prédictives	138
7.3.3	L'influence de la prévalence et de la validité intrinsèque	140
	L'influence de la prévalence	140
	L'influence de la sensibilité et de la spécificité sur les valeurs prédictives ...	142
7.4	Les combinaisons d'examens	143
7.4.1	La combinaison d'examens en parallèle	143
7.4.2	La combinaison d'examens en série.....	145
7.5	Les sources de variation dans les examens diagnostiques	147
7.5.1	La mesure de la variation: la mesure d'accord Kappa (K)	147
7.5.2	Les variations attribuables à l'examinateur	148
7.5.3	Les variations attribuables à la personne examinée	149
7.5.4	Les variations attribuables à l'examen.....	149
	En résumé	150
	Exercices	151

CHAPITRE 8

	La lecture critique des études portant sur les examens diagnostiques	152
8.1	Les critères permettant de juger de la validité d'une étude portant sur un examen diagnostique	153
8.2	La vérification de la validité interne de l'étude	155

8.2.1	La comparaison de l'examen diagnostique avec un examen de référence reconnu.....	155
8.2.2	La distribution adéquate des sujets anormaux et normaux	156
8.2.3	L'utilisation de la technique en aveugle.....	156
8.2.4	La soumission de tous les patients à l'examen de référence	157
	Un exemple de biais de vérification	157
8.3	La vérification de la reproductibilité de l'examen diagnostique	159
8.3.1	Une description explicite de la méthode utilisée qui permet de reproduire l'examen	159
	La reproductibilité de l'examen	160
8.4	La présentation des résultats de la validité de l'examen diagnostique	160
8.4.1	La validité intrinsèque et prédictive de l'examen	160
8.4.2	Le rapport de vraisemblance	161
8.5	L'utilité de l'examen diagnostique pour améliorer la qualité des soins	162
8.5.1	La vérification des avantages de l'examen diagnostique proposé	162
8.5.2	L'applicabilité des résultats de l'examen diagnostique	162
	En résumé	163
	Exercices	164
	Annexes	164

CHAPITRE 9

	La lecture critique des études descriptives: études de populations	172
9.1	L'utilité des études descriptives de populations	174
9.1.1	Les avantages d'une étude descriptive de populations.....	174
9.1.2	Les limites d'une étude descriptive de populations.....	175
9.2	Les types d'études descriptives de populations	177
9.2.1	L'étude descriptive transversale	177
9.2.2	L'étude descriptive longitudinale	178

9.2.3	L'étude de tendance.....	179
	Les effets de l'âge, de période et de cohorte	180
9.2.4	L'étude écologique.....	183
9.3	Les exigences de publication	184
9.3.1	Le titre et le résumé.....	185
9.3.2	L'introduction	186
9.3.3	La méthode	187
9.3.4	Les résultats	189
9.3.5	La discussion	190
9.3.6	La conclusion.....	191
9.3.7	Autre information.....	192
9.4	La surveillance de la santé des populations et de ses déterminants	192
9.4.1	Une description de la surveillance	192
9.4.2	Les objets de surveillance.....	194
9.4.3	L'utilité de la surveillance	194
9.4.4	Des exemples de systèmes de surveillance	195
9.4.5	Les enjeux éthiques des études descriptives de populations et de la surveillance	196
En résumé	197
Exercices	197

CHAPITRE 10

La normalité statistique et la normalité clinique	198
10.1 La notion de probabilité	199
10.1.1 La probabilité objective: fréquentiste, <i>a priori</i>	200
10.1.2 Les propriétés d'une probabilité objective	201
10.2 Les notions de population et d'échantillon	202
10.2.1 La population	202
La notion de paramètre	202
10.2.2 L'échantillon	203
La notion de statistique	205
10.3 L'inférence	206
10.3.1 Le test d'hypothèse.....	207
10.3.2 L'intervalle de confiance	207

10.3.3 La distribution de probabilités	207
La distribution normale	208
La distribution normale centrée et réduite	209
Le théorème de la limite centrale	212
10.4 La normalité statistique et la normalité clinique	216
10.4.1 La normalité statistique.....	216
10.4.2 La normalité clinique	216
En résumé	217
Exercices	217

CHAPITRE 11

La signification statistique et la signification clinique	218
11.1 La signification statistique	220
11.1.1 L'interprétation des résultats de l'inférence statistique	220
11.1.2 Les limites de l'échantillonnage	221
11.2 La réalisation d'un test d'hypothèse	224
11.2.1 L'hypothèse nulle et l'hypothèse alternative	224
11.2.2 La valeur p	226
11.2.3 Le seuil de signification statistique (α)	227
11.2.4 L'interprétation du résultat du test d'hypothèse	227
11.2.5 L'erreur de type II.....	228
11.2.6 La puissance statistique.....	230
11.3 L'intervalle de confiance	231
11.3.1 La vérification d'une hypothèse à l'aide d'un intervalle de confiance	232
La comparaison d'intervalles de confiance	232
L'intervalle de confiance et les mesures d'association	233
11.4 La signification statistique et la signification clinique	234
En résumé	236
Exercices	237
Annexes	237

CHAPITRE 12

La vision englobante de l'analyse multivariée	242
12.1 L'utilité de l'analyse univariée	243
12.1.1 La régression linéaire	244
12.2 L'utilité de l'analyse multivariée	245
12.2.1 Un rappel sur le biais de confusion	247
12.2.2 Le contrôle des biais de confusion	248
12.2.3 La modification de la mesure d'association	248
12.3 Les méthodes d'analyse multivariée	250
12.3.1 La régression linéaire multiple	251
12.3.2 La régression logistique.....	253
12.3.3 La régression à effet proportionnel ou régression de Cox.....	255
En résumé	257
Exercices	257

CHAPITRE 13

L'analyse critique des études expérimentales	258
13.1 Les facteurs susceptibles d'expliquer la réponse à un traitement ou à une intervention	259
13.1.1 L'histoire naturelle de la maladie	259
13.1.2 L'effet placebo.....	260
13.1.3 L'effet de l'attention accordée au sujet	260
13.1.4 L'effet réel d'un traitement ou d'une intervention	261
13.2 Qu'est-ce qu'une étude expérimentale?	262
13.2.1 La définition et les étapes de la démarche.....	262
13.2.2 La sélection des sujets	263
13.2.3 La répartition des sujets par hasardisation.....	265
13.2.4 Les considérations éthiques particulières aux essais cliniques	266
13.2.5 Deux plans d'études expérimentales: en parallèle et en croisé.....	266

13.3 L'analyse critique d'une étude expérimentale	269
13.3.1 L'introduction	273
Les objectifs de l'étude	273
13.3.2 La méthode et les résultats.....	274
L'évaluation de la validité interne de l'étude	274
La prévention et le contrôle des biais de sélection	275
La prévention des biais d'information ...	277
La prévention des biais de confusion ..	278
L'évaluation de la validité externe de l'étude: la précision de l'étude	281
La présentation du plan d'analyse	281
13.3.3 La discussion et la conclusion	282
L'évaluation de la validité interne de l'étude	282
L'évaluation de la taille de l'effet et sa précision	283
L'évaluation de la validité externe de l'étude	283
13.3.4 La déclaration financière	284
En résumé	284
Exercices	285

CHAPITRE 14

L'analyse critique des études à visée étiologique: études non expérimentales cas témoins et de cohorte(s)	286
14.1 L'élaboration des études à visée étiologique	288
14.1.1 Les particularités des études cas-témoins et de cohorte(s)	288
Les études cas-témoins	288
Les études de cohorte(s).....	289
14.1.2 L'exemple du café et du cancer du pancréas	290
14.2 L'analyse critique des études cas-témoins et de cohorte(s)	291
14.2.1 L'analyse critique d'une étude cas-témoins	294
Le titre et le résumé	294
L'introduction.....	295

La méthode	295	16.1 Le rôle du personnel soignant en prévention clinique	345
Les résultats	307	16.2 L'évolution des pratiques cliniques préventives	346
La discussion et la conclusion	308	16.2.1 La méthode du U.S. Preventive Services Task Force	347
Autre information	309	L'évaluation du fardeau de souffrance	347
14.2.2 L'analyse critique d'une étude de cohorte(s)	309	L'évaluation de l'efficacité de l'intervention préventive	347
Le titre et le résumé	309	Les recommandations	348
L'introduction	310	16.2.2 La stratégie d'utilisation	350
La méthode	310	16.2.3 Un retour sur le cas de Jeanne	351
Les résultats	313	16.3 La prévention et le cancer du col utérin	351
La discussion et la conclusion	314	16.3.1 Les facteurs de risque et l'histoire naturelle de la maladie	351
14.2.3 La recommandation clinique à la suite de l'analyse critique	317	16.3.2 Le dépistage	353
Note sur l'évolution de la recherche avec les devis d'observation	317	Une définition	353
14.3 Les considérations éthiques	317	Les caractéristiques des maladies dépistables	353
En résumé	318	Les caractéristiques des examens de dépistage	354
Exercices	318	Les conséquences du dépistage	356
		Le biais de durée et le biais de dépassement	357
		16.3.3 Les considérations éthiques liées au dépistage	359
		16.4 La lecture critique d'une étude portant sur l'évaluation de l'efficacité du dépistage	359
		16.4.1 La prise en compte des biais potentiels de durée et de dépassement	360
		16.4.2 La description détaillée des avantages et des inconvénients du dépistage	361
		16.4.3 La discussion sur la puissance statistique de l'étude	361
		En résumé	362
		Exercices	362
		Notes et références	364
		Corrigé	376
		Glossaire	377
		Index	389

CHAPITRE 15**La lecture critique des articles
de synthèse systématique et des
méta-analyses**

320

15.1 Les caractéristiques des différents
types d'articles de synthèse

322

15.2 Les avantages et les limites des
articles de synthèse systématique

324

15.3 Les étapes de réalisation d'une
synthèse systématique

326

15.3.1 L'interprétation des résultats

330

15.4 La critique des articles de synthèse systé-
matique à l'aide de critères objectifs

332

15.5 La modification de sa pratique à la suite
de la lecture d'un article de synthèse
systématique

341

15.6 La Collaboration Cochrane

342

En résumé

343

Exercices

343

CHAPITRE 16

**La prévention en milieu
clinique**

344

Un tour d'horizon de l'épidémiologie appliquée

Collaboration à la 3^e édition :

André Simpson

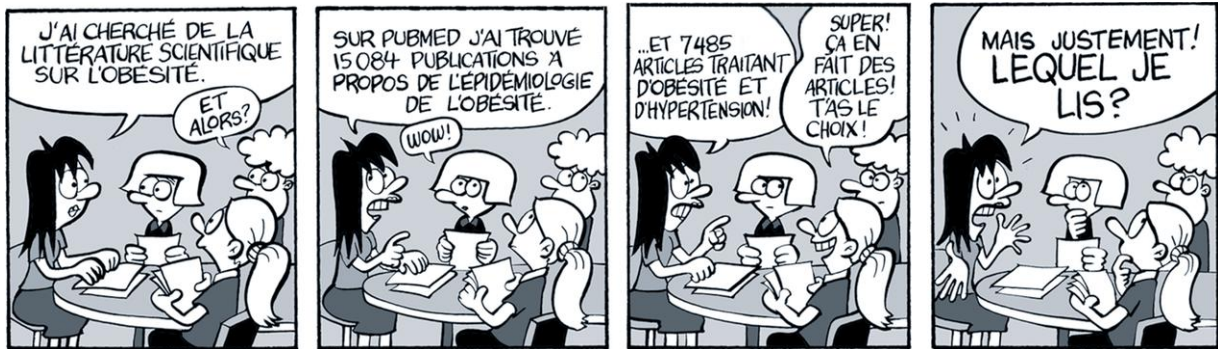
Clément Beaucage

Au terme de ce chapitre, vous pourrez :

- situer l'univers de l'épidémiologie dans votre domaine d'étude ou d'activités professionnelles ;
- expliquer l'importance de la lecture critique de la littérature en sciences de la santé.

Objectifs spécifiques	Contenu
<p>1.1 Expliquer en quoi les concepts et les méthodes en épidémiologie sont utiles dans son domaine d'étude ou d'activités professionnelles.</p>	<p>1.1 L'apport de l'épidémiologie</p> <p>1.1.1 L'épidémiologie selon une perspective populationnelle</p> <p>1.1.2 L'épidémiologie selon une perspective clinique</p> <p>1.1.3 L'épidémiologie selon une perspective interdisciplinaire</p>
<p>1.2 Décrire les grands principes éthiques qui encadrent les activités de recherche et d'intervention faisant appel à l'épidémiologie.</p>	<p>1.2 Les principes éthiques de la recherche et de l'intervention en santé</p>
<p>1.3 Expliquer les concepts et les principes de base de la lecture critique de la littérature en sciences de la santé.</p>	<p>1.3 La lecture critique de la littérature en sciences de la santé</p> <p>1.3.1 Les habiletés requises pour la lecture critique</p> <p>1.3.2 La structure générale d'une publication scientifique</p> <p>1.3.3 Un cadre de référence pour la lecture critique</p>





Mise en situation

Dans le cadre de leur rencontre interdisciplinaire hebdomadaire à la polyclinique Milo, Lisa, une omnipraticienne, Danielle, une infirmière, et Marie-Louise, une pharmacienne, discutent du fait que le nombre de personnes obèses ou ayant un surplus de poids qui se présentent à la polyclinique ne cesse d'augmenter. Quelles en seront les conséquences? Comment peut-on aborder cette question avec les patients? Bien que les causes, les répercussions sur la santé et la fréquence du surplus de poids et de l'obésité dans la population soient relativement bien documentées, les interventions efficaces en la matière font appel à une multitude de disciplines complémentaires, à savoir la médecine, les sciences infirmières, la pharmacie, la nutrition, la kinésiologie, etc.

Lisa informe ses collègues qu'elle a procédé à une recherche à l'aide d'une base de données accessible sur Internet, *PubMed*, afin de trouver de la littérature scientifique sur l'obésité. Cette seule base de données recensait 144 938 publications scientifiques en anglais ou en français ayant abordé la question du fardeau de l'obésité au cours des 10 dernières années et 17 593 ayant, pendant cette même période, traité de l'obésité et de l'hypertension. Devant une telle abondance de

littérature scientifique, que doit-on lire? Comment obtenir une information de qualité qui assure l'adoption des meilleures pratiques cliniques?

Pertinence clinique

« Il a été démontré que la lecture demeure l'activité de formation continue privilégiée par les médecins pour maintenir à jour leur compétence clinique. Ces mêmes médecins considèrent que la lecture est la méthode qui contribue le plus à modifier leur pratique¹. » Il en est de même pour les autres professionnels de la santé. L'accélération continue de la création et de la diffusion de nouvelles connaissances et technologies rend nécessaire le développement de compétences permettant une utilisation efficace et judicieuse de la littérature scientifique dans le domaine de la santé. Cela s'applique également au recours à des bases de données documentaires. À cette fin, la lecture critique de la littérature en sciences de la santé constitue un outil essentiel s'appuyant sur une connaissance fonctionnelle des concepts et des méthodes en épidémiologie. Elle sera utile pour bien circonscrire la recherche d'articles scientifiques et pour juger de la qualité et de la validité de ces publications.

1.1 L'apport de l'épidémiologie

L'**épidémiologie** est la science qui s'intéresse à la mesure de la fréquence et de la répartition des maladies, des décès et de leurs causes dans les populations. Elle a pour objet l'effet et l'interaction des facteurs de risques individuels, collectifs et

environnementaux sur la santé. Elle se penche aussi sur la mesure de l'efficacité des interventions sur la santé des individus en milieu clinique et sur celle des populations.

Traditionnellement, l'épidémiologie a été définie comme une science quantitative ayant recours essentiellement aux méthodes de la statistique et de la biostatistique. Aujourd'hui, l'épidémiologiste fait aussi appel aux méthodes qualitatives issues notamment de l'ethnographie, de l'anthropologie, de la sociologie et des sciences cliniques pour caractériser et interpréter les phénomènes d'intérêt. Ces approches quantitatives et qualitatives en épidémiologie sont complémentaires et permettent une mesure qui prend davantage en compte l'ensemble des dimensions et des déterminants de la santé des populations. On peut donc situer l'apport de l'épidémiologie selon trois perspectives majeures, soit :

1. une perspective populationnelle, c'est-à-dire portant sur l'étude des déterminants de la santé des populations, de la fréquence et de la répartition des maladies et des décès, des inégalités sociales et de santé, ainsi que sur l'étude et l'application des moyens de prévention et de contrôle de la propagation des maladies et des décès dans la population ;
2. une perspective clinique, c'est-à-dire portant sur l'étude des causes et de l'histoire naturelle des maladies et des décès, de l'efficacité des stratégies diagnostiques, des traitements et des approches cliniques préventives, et ce, dans le but d'assurer aux cliniciens l'information qui leur permettra de prendre les meilleures décisions et d'adopter les pratiques cliniques reconnues les plus efficaces ;
3. une perspective interdisciplinaire, c'est-à-dire portant sur tous les domaines des sciences de la santé.

Ces trois perspectives sont étroitement liées et complémentaires. La maladie et sa prévention ou son traitement ne sont pas le fait d'un seul individu malade ni d'un seul thérapeute ; elles constituent également un phénomène de population et de société.

1.1.1 L'épidémiologie selon une perspective populationnelle

La connaissance et la surveillance de la santé des populations et de ses déterminants constituent un enjeu important dans les sociétés actuelles. Dans plusieurs pays, la déclaration de certaines maladies afin d'en permettre la surveillance constitue une obligation légale. Dans la même optique, l'article 3 du *Règlement sanitaire international* de l'Organisation mondiale de la Santé paru en 2005² prévoit spécifiquement un volet de surveillance et de déclaration relativement au risque de propagation internationale de certaines maladies. Les concepts et les méthodes en épidémiologie constituent les outils privilégiés qui permettent de répondre à ce besoin. La mise en application de ces concepts et méthodes permet notamment :

- de mesurer la fréquence et la répartition des maladies et des décès dans les populations ;
- d'établir les causes potentielles des maladies et des décès ;
- de déterminer les stratégies et les actions de prévention et de protection de la santé des personnes et d'en mesurer l'efficacité ;

- d'anticiper la fréquence et l'évolution des principaux problèmes de santé dans les populations;
- de surveiller l'éclosion de maladies contagieuses dans le but d'en maîtriser la propagation.

Cette connaissance de la santé des populations est essentielle pour appuyer le travail des acteurs des services de santé sur la base d'informations valides qui assurent une compréhension adéquate de la réalité des individus dans leur milieu de vie. Elle améliore l'adéquation entre les services de santé, la nature des ressources cliniques accessibles et les besoins réels des individus et des populations.

La prévention de la maladie et les décès

Éviter l'apparition des facteurs de risque ou réduire l'exposition aux facteurs de risque susceptibles de provoquer les maladies et les décès constitue une stratégie de prévention. À cette fin, l'épidémiologie propose des concepts et des méthodes scientifiques qui permettent, par une approche quantitative, de cerner ces facteurs de risque et de mesurer leur impact sur la fréquence et la répartition des maladies et des décès.

La sensibilisation aux saines habitudes de vie visant à améliorer la santé des populations s'appuie également sur l'apport de l'épidémiologie, laquelle permet de découvrir les facteurs de risque sur lesquels il est possible d'agir. Les campagnes de réduction du tabagisme de même que la promotion de l'activité physique et d'une saine alimentation en sont des exemples concrets. Ainsi l'épidémiologie sera utile pour soutenir l'intervention visant à éviter que la maladie apparaisse dans une population.

L'intervention précoce et l'aide aux personnes atteintes

Dans le cas des maladies dont les causes demeurent inconnues ou pour lesquelles on ne peut agir afin de réduire l'exposition des personnes aux facteurs de risque, l'épidémiologie offre la capacité de détecter le plus tôt possible les atteintes qui en découlent. Elle permet en outre de les contrôler et d'en réduire les conséquences, ainsi que de diminuer le nombre de décès qui en résultent. L'épidémiologie s'avère alors utile pour soutenir l'intervention visant à réduire la gravité et l'impact des maladies dans les populations. On entre alors dans l'univers de l'épidémiologie appliquée à la démarche clinique.

1.1.2 L'épidémiologie selon une perspective clinique

L'anamnèse et l'évaluation de la santé d'un individu dans le contexte d'une consultation clinique ou d'un processus de traitement reposent sur l'expérience antérieure résumée dans les études et sur la connaissance de la santé des populations et de ses déterminants. Les observations cliniques sur un cas ou une série de cas conduiront à l'étude de groupes d'individus plus vastes, ou de populations, qui aura pour but de vérifier la pertinence de l'application de ces observations cliniques à grande échelle. Dans un même ordre d'idées, les connaissances acquises en épidémiologie grâce à l'étude des populations permettront de soutenir la décision et d'améliorer la pratique clinique. Les faits observés et quantifiés à grande échelle, sur des groupes d'individus ou de patients, soutiennent le jugement et la décision cliniques relatifs à un patient.

Les connaissances acquises en épidémiologie étayent également l'interprétation adéquate d'analyses de laboratoire ou d'examens diagnostiques complémentaires lorsque l'anamnèse et l'examen physique ne suffisent pas à poser un diagnostic sûr. Elles supportent le choix d'un traitement approprié et l'amélioration de la qualité des soins notamment par la connaissance de la validité de l'examen diagnostique et par l'évaluation de l'efficacité de l'intervention préventive. L'utilisation des méthodes de l'épidémiologie appliquée au domaine clinique constitue une nécessité pour appuyer les choix diagnostiques et thérapeutiques sur des bases scientifiques³.

L'épidémiologie sera donc utile aussi pour soutenir le processus et l'intervention cliniques. Il existe ainsi une grande complémentarité entre les approches cliniques et épidémiologiques.

1.1.3 L'épidémiologie selon une perspective interdisciplinaire

L'épidémiologie interpelle l'ensemble des domaines des sciences de la santé. Les lignes qui suivent illustrent bien la complémentarité, à la polyclinique Milo, du travail interdisciplinaire de Lisa, l'omnipraticienne, de Danielle, l'infirmière, de Stéphanie, la nutritionniste, de Françoise, la kinésiologue, et de Marie-Louise, la pharmacienne. L'application des concepts et des méthodes en épidémiologie dans leur domaine respectif assurera ainsi une approche globale des problèmes de santé.

Lisa, l'omnipraticienne

Lisa reçoit une de ses patientes, Geneviève, qui présente à la fois un problème d'obésité et une pression artérielle très élevée. Elle se souvient que les études épidémiologiques portant sur la pression artérielle dans des groupes d'individus ont établi un lien entre ce problème et l'obésité. Elle décèle par la suite plusieurs autres facteurs de risque susceptibles de provoquer cette situation chez Geneviève. Elle propose donc à celle-ci un traitement qu'elle estime efficace et elle répond à ses préoccupations concernant le pronostic. Sa connaissance de l'épidémiologie lui permet de lire avec intérêt et de comprendre les conclusions des études qui portent sur les facteurs de risque, l'efficacité des traitements, le pronostic, etc.

- L'analyse multivariée permet de détecter simultanément plusieurs facteurs de risque d'apparition d'une maladie dans un groupe de personnes.
- L'étude de la précision et de la validité des examens cliniques de même que les **essais cliniques et les essais thérapeutiques** permettent de juger du choix d'un traitement.
- Les concepts d'espérance de vie en bonne santé, du nombre potentiel d'années de vie perdues et de la probabilité de survie permettent d'établir un pronostic.

Les résultats des études épidémiologiques axées sur les conséquences de l'obésité ont aussi permis à Lisa de savoir que Geneviève présente, en plus des risques accrus d'hypertension artérielle, une propension au diabète de type 2, à l'insuffisance veineuse, à l'hypoventilation alvéolaire, à l'apnée du sommeil, à des dyslipidémies, à l'arthrose et

à l'insuffisance cardiaque. Tout cela peut contribuer à réduire l'espérance de vie de Geneviève ainsi que sa qualité de vie.

Conformément à l'orientation que s'est donnée la polyclinique Milo, Lisa adopte une approche interdisciplinaire et fait notamment appel à l'infirmière, à la nutritionniste, à la kinésologue et à la pharmacienne pour aborder dans sa globalité le problème d'obésité de sa patiente. Elle lui propose donc de rencontrer d'abord Danielle, infirmière à la polyclinique, qui procédera à une première évaluation et qui pourra la soutenir dans son objectif de perte de poids.

Danielle, l'infirmière

Danielle rencontre régulièrement des patientes qui, comme Geneviève, sont aux prises avec un problème d'obésité ou de surplus de poids. Elle évalue de façon systématique leur indice de masse corporelle, leur tour de taille et mesure leur pression artérielle. Elle interprète les résultats de ces mesures en s'appuyant sur les valeurs de références (ou valeurs normales) établies par l'observation de groupes d'individus dans le cadre d'études épidémiologiques.

Elle peut ainsi discuter avec ses patients de leurs problèmes de santé et des moyens d'améliorer leur condition. Sa connaissance générale de l'épidémiologie des problèmes de santé, de leur fréquence, de leur répartition, de leurs causes et de leurs conséquences sur la vie des personnes, des familles et des communautés constitue pour elle un outil précieux. Cela lui permet d'améliorer la pertinence de ses interventions, d'influencer positivement les comportements de ses patients, de situer leur problème dans un contexte plus complet et de calmer leurs inquiétudes.

Au cours de sa rencontre avec Geneviève, Danielle évalue sa motivation à perdre du poids. Elles conviennent alors de deux démarches complémentaires : Geneviève aura une consultation avec Stéphanie, une nutritionniste qui s'est nouvellement jointe à la polyclinique Milo, et une référence en kinésiologie, dans un centre médico-sportif situé près de la polyclinique.

Stéphanie, la nutritionniste

Stéphanie connaît la composition des aliments. Grâce à l'application des concepts et des méthodes en épidémiologie dans ce domaine, elle est également au courant de leurs effets sur la santé des personnes. Cette connaissance lui permet de prévenir, d'anticiper et de prendre en charge les maladies liées à la nutrition telles que le diabète, l'obésité, les pathologies cardiovasculaires et les maladies digestives. S'appuyant sur ce savoir et sur sa capacité à se tenir à jour grâce à la lecture de la littérature en sciences de la santé, Stéphanie est en mesure de conseiller et d'orienter Geneviève dans ses choix alimentaires au regard d'objectifs communs, soit une perte de poids durable et la prévention des pathologies cardiovasculaires.

Françoise, la kinésologue

La semaine suivante, Geneviève a sa première rencontre en kinésiologie au centre médico-sportif. Elle voit alors Françoise, kinésologue nouvellement diplômée, qui

discute avec elle de ses objectifs de perte de poids et de remise en forme, ainsi que du temps qu'elle veut bien y consacrer. Françoise évalue la condition physique et les habitudes d'activité physique de Geneviève. Elle prend différentes mesures anthropométriques et elle évalue la flexibilité, la force musculaire, l'endurance musculaire, la capacité aérobie, la capacité anaérobie, la composition corporelle et le pourcentage de gras de Geneviève. Pour interpréter ces mesures, Françoise s'appuie sur les résultats d'études réalisées auprès de groupes de personnes comparables à Geneviève à l'aide des concepts et des méthodes en épidémiologie. À l'occasion de sa deuxième rencontre avec Geneviève, Françoise lui propose un programme d'exercices adapté à son état de santé et qu'elle pourra entreprendre de façon sécuritaire.

Marie-Louise, la pharmacienne

Durant ce temps, Lisa discute avec sa collègue Marie-Louise, pharmaco-épidémiologiste dans un centre hospitalier universitaire, afin d'évaluer la pertinence d'avoir recours à la médication en vue de réduire l'obésité de sa patiente.

Marie-Louise collabore aux travaux de recherche et aux rencontres interdisciplinaires de la polyclinique Milo. Elle s'intéresse au traitement pharmacologique de l'obésité. Son intérêt pour la surveillance pharmacologique l'amène à être constamment à l'affût de la mise au point de nouveaux traitements, elle est captivée par l'efficacité réelle des médicaments, mais aussi par leurs effets indésirables. Elle se préoccupe également de l'efficacité de l'utilisation des médicaments, c'est-à-dire des bienfaits réels obtenus au regard des coûts encourus. Enfin, elle se soucie des habitudes de prescription et de consommation, particulièrement de celles des personnes âgées, chez qui les problèmes d'interactions médicamenteuses sont fréquents.

À l'instar de Lisa, Marie-Louise doit composer avec l'abondance de la littérature scientifique sur le sujet: une recherche rapide dans *PubMed* lui permet ainsi de trouver plus de 6 000 articles scientifiques sur l'obésité et la médication. Marie-Louise doit alors s'appuyer sur les concepts et les modèles d'études en épidémiologie, de même que sur ses habiletés de lecture critique de la littérature en sciences de la santé, pour sélectionner les articles les plus pertinents et valides afin de soutenir avec compétence les décisions relatives au recours à certains médicaments, et ce, de façon à influencer positivement les pratiques de prescription à la polyclinique Milo.

La recherche à la polyclinique Milo

Toutes ces professionnelles de la polyclinique Milo sont également engagées dans la réalisation de projets de recherche. Leur connaissance des concepts de base en épidémiologie, des types d'études épidémiologiques, des sources d'erreurs dans les études et des enjeux éthiques touchant les sujets en question leur est essentielle pour accomplir en équipe interdisciplinaire ces activités qui les passionnent.

L'épidémiologie et l'habileté à comprendre la littérature scientifique dans ce domaine ont permis à ces professionnelles de la santé de multiples secteurs de développer un langage commun et de renforcer leurs liens de collaboration au sein de l'équipe interdisciplinaire de la polyclinique Milo.

Ce cas à la polyclinique Milo dépeint les horizons de l'épidémiologie et la pertinence de la lecture critique en santé. L'exemple suivant, celui de Vincent, microbiologiste, et de sa collègue d'études Gabrielle, médecin vétérinaire, illustre d'autres horizons de l'épidémiologie.

EXEMPLE 1.1

Vincent, microbiologiste

Vincent a terminé ses études supérieures en microbiologie. Il s'intéresse particulièrement à la lutte contre les maladies infectieuses, au soutien des programmes de santé publique visant le contrôle et la surveillance de celles-ci, à l'intervention d'urgence en cas d'épidémie et à la recherche appliquée à certaines maladies. Vincent a décroché un emploi de chercheur dans un laboratoire national de santé publique grâce à ses compétences complémentaires en épidémiologie et aux habiletés qu'il a acquises en lecture critique de la littérature en sciences de la santé. En plus de participer aux travaux d'équipes de recherche multidisciplinaires, Vincent travaillera à l'élaboration d'un système de surveillance de l'influenza et d'autres maladies infectieuses, de la vaccination et de la gestion des produits immunisants.

Gabrielle, médecin vétérinaire

Gabrielle, consœur d'études de Vincent en microbiologie, s'est orientée vers la médecine vétérinaire dans le cadre de ses études supérieures. Elle s'intéresse particulièrement à la production animale et à la santé publique, notamment à la transmission des maladies animales chez l'humain. À la fin de ses études, elle a obtenu un emploi dans le même organisme national de santé publique que Vincent. Elle travaille entre autres à la surveillance du virus du Nil occidental en effectuant le suivi des mesures de contrôle et de surveillance entomologique afin d'anticiper les risques de transmission chez l'humain. Elle est également demandée pour la surveillance et le contrôle des risques de transmission de la rage, qui touche de façon épidémique une espèce animale sauvage de la région du sud du pays. Dernièrement, Gabrielle a été heureuse d'apprendre qu'elle devait implanter dans son organisation un plan de lutte contre la pandémie d'influenza et qu'à ce titre, elle serait appelée à travailler en étroite collaboration avec son confrère d'études Vincent.

Pour le travail d'équipe

Comme on peut le constater, les compétences acquises par Gabrielle et Vincent dans le domaine de l'épidémiologie leur sont essentielles pour mener à bien ce qui les passionne, soit l'amélioration de la santé des populations humaines et animales. Les concepts de base et les modèles d'études en épidémiologie portant sur la fréquence et la répartition, les déterminants, le contrôle ainsi que l'étiologie des maladies et les causes des décès leur permettent d'œuvrer dans ce domaine au sein d'équipes multidisciplinaires tout en disposant d'un référentiel et d'un langage communs favorisant des collaborations crédibles et performantes.

1.2 Les principes éthiques de la recherche et de l'intervention en santé

La réalisation d'études épidémiologiques à partir de sources de données existantes ou s'appuyant sur une collecte de données systématique auprès d'individus de même que la participation, dans un contexte clinique, à des études impliquant l'expérimentation chez l'humain (essais cliniques, essais thérapeutiques visant à mesurer l'effet d'un traitement ou d'un médicament) soulèvent des enjeux éthiques importants. La prise en compte de ces enjeux éthiques constitue une partie intégrante de tout protocole de recherche.

Les études épidémiologiques sont régies par des règles éthiques strictes et se trouvent soumises à des contrôles précis visant à assurer la protection des personnes qui y participent (par exemple, Fonds de recherche en santé du Québec⁴, Golberg⁵, UNESCO⁶). De telles études doivent démontrer leur utilité, leur pertinence et les bienfaits qu'elles apporteront aux personnes qui y prennent part ainsi qu'aux populations que celles-ci représentent. Ces aspects éthiques ont fait l'objet d'un grand nombre d'analyses et de publications. Nous retiendrons ici à titre de référence la *Déclaration universelle sur la bioéthique et les droits de l'homme* adoptée par l'UNESCO en octobre 2005⁷. Cette déclaration résume l'essentiel des considérations éthiques traitées dans des ouvrages antérieurs tout en y ajoutant certaines dimensions plus contemporaines, comme le respect de la diversité culturelle et du pluralisme, la protection des générations futures, de même que la protection de l'environnement, de la biosphère et de la biodiversité. Le tableau 1.1 présente les principes éthiques tirés de cette déclaration.

TABLEAU 1.1 ■ Les principes éthiques régissant les études épidémiologiques

Individus	<ul style="list-style-type: none"> • Éviter la discrimination et la stigmatisation. • Respecter la vie privée et la confidentialité. • Respecter la vulnérabilité humaine et l'intégrité personnelle. • Protéger les personnes incapables d'exprimer leur consentement. • Assurer un consentement éclairé. • Assurer l'autonomie et la responsabilité individuelle. • Maximiser les effets bénéfiques et minimiser les effets nocifs. • Respecter la dignité humaine et les droits de la personne.
Société	<ul style="list-style-type: none"> • Partager les bienfaits. • Promouvoir la responsabilité sociale et sanitaire. • Promouvoir la solidarité et la coopération. • Respecter la diversité culturelle et le pluralisme. • Promouvoir l'égalité, la justice et l'équité.
Environnement	<ul style="list-style-type: none"> • Protéger l'environnement, la biosphère et la biodiversité. • Protéger les générations futures.

L'application de ces principes fait l'objet de contrôles systématiques par les comités d'éthique et les autres instances devant autoriser ou encadrer les activités de recherche. Les questions suivantes illustrent les types de contrôles réalisés⁸.

- Quels sont les risques et les effets secondaires possibles chez les patients recrutés ?
- Quels sont les inconvénients encourus par les participants (séjours et tests supplémentaires à l'hôpital, frais personnels, etc.) ?
- Quelles sont les mesures de protection des renseignements confidentiels mises en place ?
- Quels sont les principaux avantages pour les patients recrutés dans le cadre du projet ?
- Les participants sont-ils en mesure d'émettre un consentement éclairé et libre de toute contrainte ?
- Quelles sont, le cas échéant, les mesures envisagées pour protéger les droits des mineurs, des handicapés mentaux ou des personnes légalement incompétentes ?
- Quelles sont les conditions de retrait du projet pour les patients recrutés ?

1.3 La lecture critique de la littérature en sciences de la santé

Pour se révéler utiles, les connaissances scientifiques doivent être communiquées. L'expérience du clinicien, qui se construit grâce au cumul des observations réalisées auprès de ses patients, ne doit pas constituer la seule voie pour assurer une pratique clinique des plus performantes. La communication des observations scientifiques réalisées par les cliniciens de même que le partage des connaissances qui résultent de l'observation de la santé d'individus et de populations sont essentiels pour favoriser l'évolution des pratiques par l'intégration des connaissances scientifiques nouvelles.

1.3.1 Les habiletés requises pour la lecture critique

Autrefois, la communication de nouvelles connaissances se faisait principalement par des publications dans les revues scientifiques spécialisées. Depuis plusieurs années maintenant, Internet en a permis une diffusion accélérée. L'utilisation du Web pour accéder à la littérature scientifique en sciences de la santé est devenue plus fréquente que le recours aux documents imprimés. Il est donc d'autant plus essentiel de pouvoir poser un jugement critique juste sur la qualité d'une publication ou d'un résultat scientifique. C'est en effet ce jugement qui viendra appuyer les décisions dans un contexte clinique ou de santé publique. En santé publique, il permet par exemple d'évaluer la pertinence de mettre en place une activité préventive, un programme de sensibilisation, une intervention en protection de la santé ou un programme de dépistage précoce de certaines maladies. Dans un contexte de service personnalisé, ce résultat appuie l'exercice du jugement clinique et soutient l'établissement d'un diagnostic et du choix d'un traitement, de même que la décision de modifier une pratique clinique à la lumière de nouvelles connaissances.

En somme, il y a «lire» et «savoir lire». Lire un document est une chose; être en mesure de cibler les meilleurs écrits, de comprendre tout ce qui y figure et d'en dégager l'essentiel de façon à en tirer pleinement profit en est une autre. La lecture critique de la littérature en sciences de la santé dont traite le présent ouvrage vous permettra d'acquérir ces habiletés.

1.3.2 La structure générale d'une publication scientifique

La majorité des chapitres de cet ouvrage s'appuient sur des articles scientifiques qui constituent autant d'exemples de la structure générale d'une publication scientifique. Ces publications répondent à des normes et à des exigences qui en assurent l'uniformité. Elles sont structurées de façon à présenter d'abord un résumé qui couvre l'ensemble des sections suivantes : l'introduction, la méthode, les résultats, la discussion et la conclusion. Ces sections sont ensuite reprises dans le corps de l'article, que clôt une liste de références.

- **Le résumé.** L'article débute par un résumé des différentes sections de l'étude. Le résumé permet au lecteur de vérifier dès le départ si l'article correspond à ses préoccupations et à ses champs d'intérêt. Il donne les grandes lignes de l'étude réalisée et des résultats observés.
- **L'introduction.** Cette section expose les grandes lignes de l'étude réalisée et souligne de façon succincte l'intérêt qu'elle revêt en s'appuyant sur des références pertinentes. On y présente, par une revue de la littérature, un bilan de l'état des connaissances de l'objet de la recherche. L'auteur y définit son approche du sujet ainsi que les objectifs de l'étude.
- **La méthode.** La section «Méthode» décrit clairement la population à l'étude ainsi que la façon dont les sujets ont été sélectionnés, la ou les sources de données analysées, les informations analysées, etc. Cette section présente également un plan d'analyse des données. On y trouve notamment des précisions sur les méthodes statistiques utilisées.
- **Les résultats.** La section «Résultats» fournit une description des caractéristiques pertinentes des personnes observées. On y trouve aussi les résultats des mesures réalisées auprès des individus à l'étude, par exemple la fréquence et la répartition des maladies étudiées, ainsi que des décès et de leurs causes. Selon le cas, les résultats sont présentés sous forme de tableaux et de figures; les interprétations des tests statistiques y figurent également.
- **La discussion.** On trouve dans cette section une interprétation des résultats de l'étude. Les auteurs en discutent et présentent les éléments qui ont pu influencer ces résultats, par exemple des erreurs pouvant découler de la méthode utilisée, certaines caractéristiques des personnes étudiées, etc. Une comparaison avec les résultats d'autres études permet en outre de discuter des résultats observés. C'est aussi dans cette section que les limites de l'étude sont présentées.
- **La conclusion.** Cette dernière section reprend l'essentiel des résultats de l'étude. Elle fait ressortir ce qui doit être retenu en fonction des objectifs initiaux de l'étude. La conclusion permet également de suggérer des pistes de recherche complémentaire à la lumière des résultats observés.

1.3.3 Un cadre de référence pour la lecture critique

Bien que cette structure normalisée des publications scientifiques en facilite la lecture, l'abondance de la littérature dans le domaine de la santé et son accès grandement facilité par les grandes bases de données regroupant les publications scientifiques de divers domaines exigent que l'on se dote d'une méthode de travail permettant de tirer profit de cette grande quantité de connaissances nouvelles, selon une approche objective et en y consacrant un temps optimal. Durant les 15 dernières années, la lecture critique d'articles a fait l'objet d'un développement méthodologique important et de nombreuses publications. On trouve aujourd'hui des guides de lecture critique et de rédaction pour différents types de publications scientifiques du domaine de la santé. Le Réseau Equator⁹ qui vise à améliorer la qualité et la transparence de la recherche en santé regroupe différents experts ayant produit des guides de rédactions et listes de vérifications permettant une approche méthodique et rigoureuse de la littérature en sciences de la santé¹⁰.

Prenons Audet et Leclère¹¹ qui proposent, en s'appuyant principalement sur une recension des écrits, un cadre de référence des composantes de cette compétence en lecture critique des publications scientifiques en santé. Ces auteures établissent ainsi les compétences requises pour réaliser les étapes avant, pendant et après la lecture des articles scientifiques. Ces compétences sont présentées dans l'encadré 1.1, inspiré des travaux de ces auteures.

ENCADRÉ 1.1 ■ Le cadre de référence concernant la compétence à critiquer une publication scientifique en sciences de la santé

1. Avant la lecture, le professionnel choisit ce qu'il va lire. Il doit :

- maîtriser le sujet: posséder des schémas cognitifs appropriés au sujet traité;
- préciser un problème ou une préoccupation fondamentale ou clinique motivant sa lecture;
- élaborer et maîtriser des stratégies efficaces de recherche d'articles (banque de données, Internet, bibliothèque, expert ou archivage d'articles), afin de retrouver rapidement un article susceptible de résoudre son problème ou sa préoccupation.

Première décision: il décide de lire l'article sélectionné.

2. Par une première lecture rapide, il décide si c'est le bon article. Il doit :

- s'assurer d'être concerné par le sujet traité;
- s'assurer de bien connaître le contenu clinique et les concepts en biostatistique et en épidémiologie clinique utilisés dans l'article;
- vérifier la pertinence de la publication en faisant le lien entre l'étude et sa préoccupation.

Deuxième décision: cet article est pertinent.

3. Pendant la lecture, il décide s'il s'agit d'une étude valide sur le plan scientifique. Il doit :

- comprendre les concepts scientifiques en biostatistique et en épidémiologie clinique utilisés dans l'étude;



ENCADRÉ 1.1 ■ Le cadre de référence concernant la compétence à critiquer une publication scientifique en sciences de la santé (suite)

- évaluer la justesse des données :
 - distinguer un fait (méthode et résultat) d'un commentaire (introduction et discussion) ;
 - détecter les erreurs systématiques (les biais) ;
 - vérifier la qualité et la pertinence des tableaux et des figures ;
 - s'assurer de l'exhaustivité des données ;
- évaluer l'interprétation des données :
 - déterminer la force de la preuve ;
 - critiquer le choix du devis de recherche ;
 - critiquer le choix des tests statistiques utilisés ;
 - s'assurer que la généralisation respecte les limites permises par la méthodologie et par la population de référence.

Troisième décision : cette étude est valide sur le plan scientifique.

4. Après la lecture, il décide si cette étude est utile et s'il modifiera sa pratique. Il doit :

- juger de l'utilité clinique des résultats pour sa pratique en général ;
- déterminer si la réponse à sa préoccupation initiale est partielle ou complète.

Quatrième décision : cette étude est utile pour sa pratique clinique.

Il doit :

- juger de l'utilité clinique des résultats pour sa propre clientèle ;
- décider de modifier ou non sa pratique en fonction des résultats de l'étude ;
- définir concrètement les implications de cette décision sur sa démarche et ses décisions cliniques.

Cinquième décision : cette étude modifie sa pratique.

Source : Adapté de Audet, N. et H. Leclère. «Les habiletés requises pour la lecture critique en médecine : un cadre de référence issu d'une recension des écrits», *Pédagogie médicale*, 2001 vol. 2, n° 4, p. 207.

En définitive, l'approche méthodique de la littérature scientifique en santé de même qu'une bonne connaissance des concepts et des méthodes de base en épidémiologie et en biostatistique sont les conditions essentielles à la réussite scolaire de l'étudiant dans ces domaines. Elles constituent également la recette de l'actualisation des connaissances requises pour l'évolution et l'adoption des meilleures pratiques professionnelles en santé. Par exemple, une telle démarche permettra aux professionnels de la polyclinique Milo d'y voir plus clair parmi les nombreuses publications sur l'obésité et ses effets sur la santé des personnes, ainsi que de mieux cibler leurs préoccupations, leurs analyses et leurs discussions en vue d'adopter les meilleures pratiques pour servir les patients de la polyclinique. Le très grand nombre de publications constitue également un défi : nous ne pouvons pas réviser 20 000 titres afin de sélectionner les publications scientifiques les plus pertinentes à notre préoccupation. Il importe alors

de bien circonscrire notre recherche documentaire. Les principales bases de données documentaires et les bibliothèques universitaires fournissent notamment des guides en ce sens¹².

En résumé

Pour toutes les disciplines du secteur de la santé, les habiletés et les compétences en matière de lecture critique s'avèrent indispensables aux apprentissages de base, à la formation continue et à l'adoption des meilleures pratiques cliniques. L'épidémiologie permet d'aborder la santé, la maladie et le bien-être, de même que l'interaction des facteurs qui les influencent, en considérant les groupes, les familles, les populations ou les sociétés au-delà des personnes qui les composent. En plus de fournir des connaissances

fondées sur les évidences et les preuves pour soutenir la démarche clinique, les concepts et les méthodes épidémiologiques procurent aux intervenants des références et un langage communs permettant de renforcer les collaborations dans le cadre d'équipes interdisciplinaires. Dans ce domaine comme dans d'autres, la recherche soulève des questions éthiques. Le développement d'une grande sensibilité à cet égard est indispensable et doit demeurer une préoccupation constante.

Exercices



Voir la section complète des exercices sur le Web.

1. L'apport de l'épidémiologie peut se situer selon une perspective clinique, une perspective populationnelle et une perspective interdisciplinaire. Situez chacun des énoncés suivants selon l'une de ces perspectives.
 - a) Mesurer la fréquence et la répartition des maladies et des décès.
 - b) Mesurer l'efficacité d'un traitement.
 - c) Fournir au médecin, à l'infirmière et au kinésologue les données sur les interventions les plus efficaces pour réduire le risque de chute chez les personnes âgées.
 - d) Mesurer les inégalités sociales de santé.
 - e) Étudier les causes des maladies et des décès.
2. Quel principe éthique n'est pas respecté si un patient inconscient admis à l'urgence est immédiatement intégré dans un protocole de recherche visant à mesurer l'efficacité d'un nouveau traitement?
 - a) Le respect de la vulnérabilité humaine et de l'intégrité des personnes.
 - b) L'obtention d'un consentement éclairé.
 - c) Le respect de la dignité humaine et des droits de la personne.
3. Dans quelle section d'une publication scientifique trouve-t-on les informations sur les sources d'erreurs potentielles et sur les limites de l'interprétation des résultats de l'étude publiée?
 - a) Introduction.
 - b) Méthode.
 - c) Résultats.
 - d) Discussion.
 - e) Conclusion.
4. Choisissez la bonne réponse. L'approche méthodique de la littérature scientifique en santé et la lecture critique sont des incontournables compte tenu :
 - a) de la complexité grandissante des publications scientifiques sur le plan méthodologique.
 - b) du grand nombre de publications scientifiques avec lequel il faut composer et de la nécessité de s'appuyer sur les meilleures études.
 - c) de la grande diversité des publications scientifiques dans divers domaines de spécialisation.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 2

L'approche statistique de la réalité

Collaboration à la 3^e édition :

André Simpson

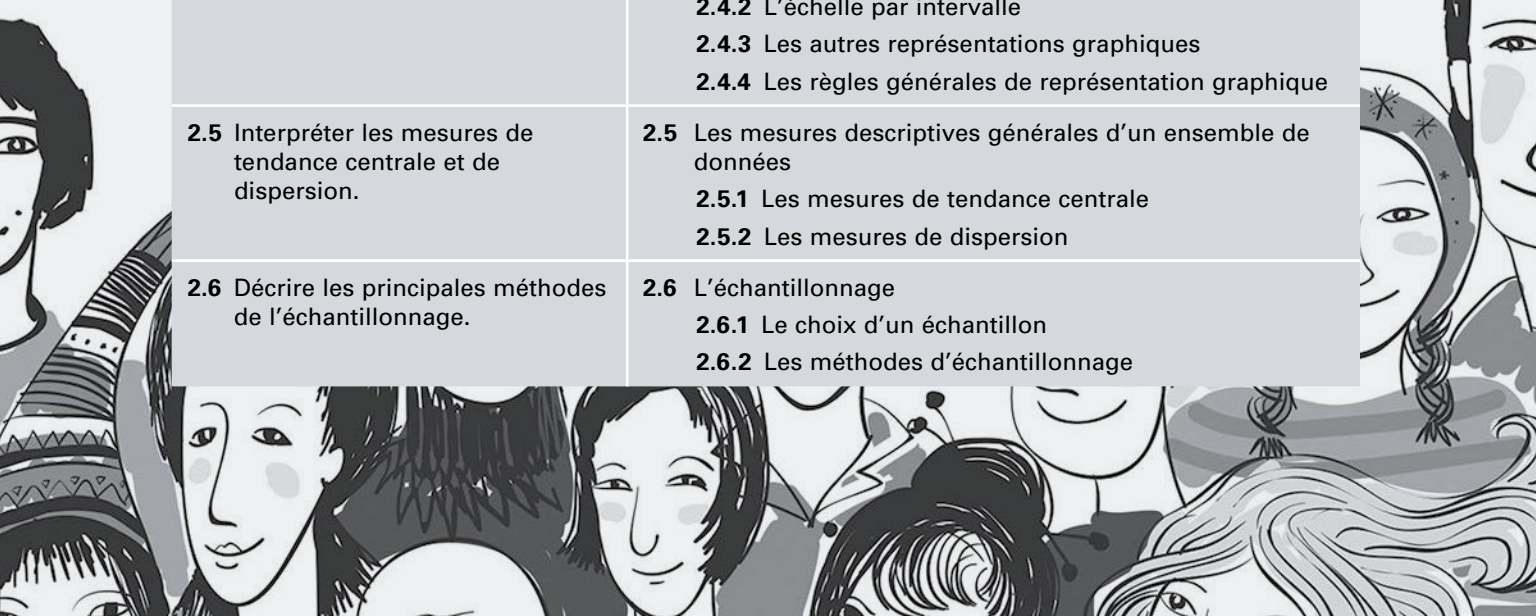
Danielle Laurin

Thomas Matukala Nkosi

Au terme de ce chapitre, vous pourrez :

- interpréter adéquatement les informations quantitatives de base présentées dans la littérature en sciences de la santé ;
- utiliser les concepts de base de l'échantillonnage.

Objectifs spécifiques	Contenu
2.1 Expliquer la notion de variable.	2.1 La notion de variable 2.1.1 Les types de variables 2.1.2 La classification des variables
2.2 Décrire les échelles de classification.	2.2 La classification des observations 2.2.1 L'échelle de classification 2.2.2 Les types d'échelles de classification 2.2.3 Le passage d'une échelle à une autre
2.3 Interpréter un tableau de fréquences.	2.3 Les distributions de fréquences
2.4 Interpréter les représentations graphiques des distributions de fréquences.	2.4 Les représentations graphiques des distributions de fréquences 2.4.1 L'échelle nominale ou ordinale 2.4.2 L'échelle par intervalle 2.4.3 Les autres représentations graphiques 2.4.4 Les règles générales de représentation graphique
2.5 Interpréter les mesures de tendance centrale et de dispersion.	2.5 Les mesures descriptives générales d'un ensemble de données 2.5.1 Les mesures de tendance centrale 2.5.2 Les mesures de dispersion
2.6 Décrire les principales méthodes de l'échantillonnage.	2.6 L'échantillonnage 2.6.1 Le choix d'un échantillon 2.6.2 Les méthodes d'échantillonnage





Mise en situation

Anne-Marie accompagne son fils Yako à la polyclinique Milo où travaillent Danielle, l'infirmière, et Lisa, l'omnipraticienne. C'est la mi-septembre, et l'école vient de commencer. Anne-Marie s'inquiète du fait que Yako est le plus petit de sa classe. Danielle le mesure et le pèse, puis note les observations.

Yako a 5 ans et 3 mois : il mesure 105 cm et pèse 17 kg. Son état de santé général est excellent. À la lecture de son carnet de santé, Danielle constate que, depuis sa naissance, il s'est toujours situé dans le groupe des 10% plus petits quant au poids et à la taille.

Pertinence clinique

En clinique, le professionnel fait des observations sur une personne. Pour en tirer une conclusion, il les compare toujours à la « normale ». Ce que l'on appelle « valeur normale », ou « valeur de référence », provient d'un résumé de milliers d'observations antérieures. Comment cette connaissance du « normal » est-elle acquise ? Comment chaque personne peut-elle contribuer à enrichir l'expérience collective à partir de la sienne ? Pour que chacun puisse apporter sa contribution, il est nécessaire que tous partagent un même langage pour exprimer la réalité collective, et c'est ce que proposent la statistique et l'épidémiologie.

2.1 La notion de variable

Yako mesure 105 cm, et son copain Frédéric, 119 cm. Le groupe sanguin de sa mère est A, et celui de sa gardienne, Jeanne, est B. La taille des garçons et des filles composant une famille de 8 enfants, dont l'âge s'échelonne de 3 ans à 18 ans, varie de 100 cm à 180 cm. Un hôpital engage des médecins, des infirmières, des techniciennes de laboratoire, etc. Dans le domaine de l'épidémiologie, la taille, le groupe sanguin, le sexe, l'âge et le type de travail constituent des variables. On appelle **variable** toute caractéristique susceptible d'être différente selon les personnes, les lieux ou le temps.

EXEMPLE 2.1

- L'âge, le sexe, le groupe sanguin, la tension artérielle et le nombre de lits par hôpital sont des variables.

À la notion de variable s'ajoute celle de **valeur**. Tout état que prend la variable étudiée est une valeur.

EXEMPLE 2.2

- La variable *sexe* a deux valeurs : masculin, féminin.
- La variable *nombre de lits par hôpital* a un nombre déterminé de valeurs : 0, 1, 2, ..., n lits.
- La variable *groupe sanguin* a quatre valeurs : A, B, AB, O.
- La variable *taille* a un nombre indéterminé de valeurs : 125 cm, 125,5 cm, 150,531 cm, etc.

2.1.1 Les types de variables

En épidémiologie, les variables peuvent être regroupées suivant trois dimensions qui permettent de caractériser la maladie ou la santé : les variables de personnes, de lieux et de temps.

Les variables de personnes

Les **variables de personnes** renvoient aux attributs anatomiques, physiologiques, sociaux, économiques ou culturels. Les plus fréquemment considérées sont l'âge, le sexe, l'état civil, les habitudes de vie, l'occupation et le statut socioéconomique. Les variables de personnes permettent de répondre à la question : « De qui parle-t-on ? »

EXEMPLE 2.3

Yako est un garçon de 5 ans et 3 mois. Il pèse 17 kg et mesure 105 cm.

Les variables de lieux

La fréquence d'une maladie peut varier selon le pays, la région ou la situation géographique de la population (zone urbaine ou rurale). Les **variables de lieux** permettent de répondre à la question : « Où se déroulent les événements dont on parle ? »

EXEMPLE 2.4

- La région de Lanaudière a connu une épidémie de rougeole.
- La mère de Yako travaille dans une manufacture de chaussures.

Les variables de temps

De façon générale, la fréquence d'une maladie varie avec le temps. La durée est une caractéristique de la maladie qui permet de marquer sa gravité et son évolution. Le temps est donc un élément nécessaire à la définition des mesures épidémiologiques et une composante de base des concepts de cause. Les **variables de temps** permettent de répondre à la question : « Quand l'événement est-il survenu ? »

EXEMPLE 2.5

- Au début de l'année scolaire, Anne-Marie s'inquiète de la taille de Yako.
- Le cancer du poumon a fortement augmenté chez les femmes depuis 20 ans.
- Trois heures après son repas, Jeanne a éprouvé de sérieuses crampes abdominales.

2.1.2 La classification des variables

Les variables ne sont pas toutes de même nature. Elles se distinguent d'abord selon que leurs valeurs sont de type numérique ou non : le poids des enfants est une variable

de type numérique, alors que la couleur de leurs vêtements ne l'est pas. Le fait que ces valeurs soient isolées ou non les unes par rapport aux autres constitue un autre facteur distinctif. Ainsi, entre une classe de 20 enfants et une autre de 21 enfants, il n'y a pas de valeur intermédiaire possible; le nombre d'enfants dans une classe est donc une variable dont les valeurs sont isolées. Par contre, entre les kilogrammes, il y a toujours les grammes, entre les grammes, les milligrammes, entre les milligrammes, les microgrammes, etc.; les valeurs du poids ne sont donc pas isolées. Ces distinctions entraînent la classification des variables en variables qualitatives ou quantitatives; ces dernières peuvent être discrètes ou continues.

Les variables quantitatives

La valeur de la **variable quantitative** est numérique. On partage les valeurs numériques en valeurs discrètes et en valeurs continues.

Une variable quantitative est discrète lorsque ses valeurs sont des quantités isolées, séparées les unes des autres. Les valeurs d'une telle variable sont obtenues par dénombrement.

EXEMPLE 2.6

Anne-Marie a trois enfants. La variable *nombre d'enfants par famille* peut prendre les valeurs 0, 1, 2, 3, etc. Entre les valeurs observables 2 et 3, il n'y a pas de nombre intermédiaire. Une famille peut compter 2 ou 3 enfants, mais pas 2,6.

Ainsi, une **variable discrète** ne peut être exprimée par une fraction. À l'inverse, une **variable** est **continue** lorsque ses valeurs sont des quantités qui peuvent s'exprimer par une fraction.

EXEMPLE 2.7

La taille de Yako augmentera de 105 à 109 cm d'ici quelques mois. Ce faisant, elle passera nécessairement par toutes les valeurs intermédiaires possibles qui s'expriment par une fraction, que ce soit 106,5 cm ou 108,99999 cm.

Les variables qualitatives

La valeur de la **variable qualitative** ou **catégorielle** correspond à des qualités, des attributs. Ainsi en est-il du sexe (masculin, féminin) et du groupe sanguin (A, B, AB, O).

La variable qualitative est, par convention, discrète. En effet, même si l'on peut observer une continuité dans le ton d'une couleur (de très pâle à très foncé), on conviendra par exemple de les catégoriser en cinq groupes comme «très pâle», «pâle», «moyen», «foncé» et «très foncé».

Le choix des instruments de description statistique et de mesure d'une variable dépend de la nature de celle-ci. Il importe donc de savoir discerner si une variable est qualitative ou quantitative et, dans ce dernier cas, si elle est discrète ou continue. Ainsi, pour une variable quantitative comme le nombre d'enfants par famille ou la pression artérielle systolique, on calculera la moyenne arithmétique. Pour une valeur qualitative, on calculera plutôt une proportion, comme la proportion d'enfants du groupe sanguin A dans la classe de Yako.

2.2 La classification des observations

Que l'on procède à une collecte de renseignements sur une série de variables ou que l'on s'appuie sur une ou plusieurs sources d'informations, inévitablement, on se retrouve avec un ensemble de données comme celles présentées dans le tableau 2.1.

TABLEAU 2.1 ■ Un exemple d'un ensemble de données

N° de dossier	Identifiant	Âge (années)	Poids (kg)	Taille (mètres)	Occupation	Lieu de résidence	Groupe sanguin
1	Yako	5	17	1,15	Écolier	Québec	A
2	Anne-Marie	26	52	1,71	Enseignante	Québec	B
3	Jeanne	23	55	1,73	Éducatrice	Québec	B
4	Denis	48	68	1,81	Gestionnaire	Montréal	A
(...)	–	–	–	–	–	–	–
580	Yves	45	74	1,91	Biologiste	Lévis	AB

L'ensemble de données peut contenir, pour chaque variable, autant de valeurs qu'il y a d'événements, par exemple d'individus, observés. Le défi consiste alors à produire une information utile et intelligible, souvent à partir d'un grand nombre de valeurs observées pour une même variable.

La première étape de ce travail d'organisation des valeurs observées réside dans leur classement pour chaque variable de l'ensemble de données. Ainsi, pour chaque variable, on dénombre les valeurs identiques.

Une autre stratégie consiste à regrouper ces valeurs dans des classes, puis à les dénombrer à l'intérieur de chaque classe.

EXEMPLE 2.8 (voir le tableau 2.1)

- Anne-Marie a 26 ans. Elle appartient à la classe d'âge «25-29 ans», tout comme les 30 autres personnes qui ont de 25 à 29 ans dans l'ensemble de données. On dénombre donc 31 observations dans cette classe d'âge.
- Jeanne a 23 ans. Elle appartient à la classe d'âge «20-24 ans». Il en est de même pour les 15 autres personnes qui ont de 20 à 24 ans dans l'ensemble de données. On dénombre donc 16 observations dans cette classe d'âge.

2.2.1 L'échelle de classification

Pour une variable donnée, l'ensemble des classes constitue une échelle de classification.

EXEMPLE 2.9

Les quatre classes A, B, AB, O (*voir le tableau 2.1*) constituent une échelle de classification pour la variable *groupe sanguin*.

Une échelle de classification doit permettre de distribuer toutes les observations, chacune ne pouvant appartenir qu'à une seule catégorie. Pour qu'un classement des observations soit juste, les classes qui constituent l'échelle doivent nécessairement satisfaire à deux conditions.

D'une part, elles doivent être mutuellement exclusives : chaque individu ou chaque observation de la variable ne peut appartenir qu'à une seule classe.

EXEMPLE 2.10

Les classes d'âge «1-5 ans», «5-15 ans», «15-25 ans» et «25 ans et plus» ne sont pas mutuellement exclusives, car les individus de 5 ans, 15 ans et 25 ans appartiennent à plus d'une classe. Par contre, les classes «1-4 ans», «5-14 ans», «15-24 ans» et «25 ans et plus» le sont.

D'autre part, elles doivent être collectivement exhaustives : chaque individu ou chaque observation de la variable doit appartenir à une classe.

EXEMPLE 2.11

- Les classes d'âge «1-4 ans», «5-14 ans», «15-24 ans» et «25 ans et plus» ne sont pas collectivement exhaustives si un enfant âgé de moins de 1 an se retrouve dans l'ensemble de données. Il faut alors ajouter la classe d'âge «Moins de 1 an» pour en faire des classes collectivement exhaustives.
- Si l'on choisissait de classer les personnes selon les groupes sanguins A ou O, on n'obtiendrait pas une échelle comportant des classes collectivement exhaustives puisque Jeanne, qui est du groupe B (*voir le tableau 2.1*), ne pourrait être classée. Il faudrait donc ajouter à l'échelle les classes «B», «AB» – et même prévoir au besoin une catégorie pour les données manquantes.

2.2.2 Les types d'échelles de classification

En distinguant les variables qualitatives des variables quantitatives et les variables discrètes des variables continues, on peut répartir les échelles de classification selon trois types.

L'échelle nominale

Dans une **échelle nominale**, les classes ne sont que nommées. Le tableau 2.2A présente quelques exemples de variables dont l'échelle est nominale.

L'échelle ordinale

Dans une **échelle ordinale**, les classes sont nommées et ordonnées, allant de la plus petite valeur à la plus grande ou de la plus grande valeur à la plus petite. Le tableau 2.2B présente quelques exemples de variables dont l'échelle est ordinale.

L'échelle par intervalle

Dans une **échelle par intervalle**, les classes sont nommées et ordonnées; il existe de plus une relation de distance entre les valeurs. Le tableau 2.2C présente un exemple de variable dont l'échelle est par intervalle.

TABLEAU 2.2 ■ Des exemples de variables selon l'échelle

A. Une échelle nominale		B. Une échelle ordinale		C. Une échelle par intervalle	
Variable	Classes	Variable	Classes	Variable	Classes
Sexe	<ul style="list-style-type: none"> • Masculin • Féminin 	Évolution de l'état de santé	<ul style="list-style-type: none"> • Amélioration • Stabilité • Détérioration 	Âge	• 1 – 4
Groupe sanguin	• A, B, O, AB				• 5 – 9
Diagnostic	<ul style="list-style-type: none"> • Cystite • Pyélonéphrite • Autres 	Degré de satisfaction	<ul style="list-style-type: none"> • Insatisfait • Peu satisfait • Satisfait • Très satisfait 		• 10 – 14
					• 15 – 19
					• 20 – 24
					• 25 – 29
				• 30 – 34	

EXEMPLE 2.12

Yako appartient à la classe d'âge «5-9 ans», Jeanne, à la classe «20-24 ans», et Anne-Marie, à la classe «25-29 ans». On peut donc estimer que la distance ou l'écart entre les enfants de l'âge de Yako et les adultes de l'âge de sa mère Anne-Marie est approximativement de 20 ans. Par contre, cet écart pour les personnes de la tranche d'âge de Jeanne est approximativement de 5 ans.

2.2.3 Le passage d'une échelle à une autre

Il existe une hiérarchie dans les échelles: il est possible de passer de l'échelle par intervalle à l'échelle ordinale, puis à l'échelle nominale. Chaque regroupement des données pour effectuer le passage d'une échelle à l'autre entraîne une perte d'information. Il s'agit donc, en quelque sorte, d'une voie à sens unique. On aura noté que la même perte d'information survient quand on passe des données brutes à un regroupement en classes.

EXEMPLE 2.13

Stéphanie, nutritionniste à la polyclinique Milo, passe en revue les dossiers de ses patients. Elle peut noter, dossier après dossier, la cholestérolémie de chacun d'entre eux. Toutefois, si elle veut présenter un tableau d'ensemble à ses collègues de la polyclinique, elle doit classer les observations.

Si elle utilise une échelle par intervalle, elle perdra l'information sur chacun des patients, mais pourra connaître le nombre de ceux dont la cholestérolémie se situe de 2,8 mmol/l à 3,3 mmol/l, de 3,4 mmol/l à 3,9 mmol/l, ..., de 5,2 mmol/l à 5,9 mmol/l, etc.

Si elle désire présenter une information plus facile à saisir, elle peut construire une échelle ordinale en définissant les classes suivantes: cholestérolémie basse,

EXEMPLE 2.13 (suite)

cholestérolémie peu élevée, cholestérolémie élevée et cholestérolémie très élevée. Ainsi, les classes «2,8-3,3 mmol/l» et «3,4-3,9 mmol/l» pourraient être regroupées dans la classe «cholestérolémie basse»; les classes «4,0-4,4 mmol/l», «4,5-4,9 mmol/l» et «5,0-5,4 mmol/l» pourraient être regroupées dans la classe «cholestérolémie peu élevée», et ainsi de suite.

Enfin, pour simplifier davantage, Stéphanie peut broser un tableau de la cholestérolémie de sa clientèle en supprimant toute notion d'ordre et en créant deux catégories: cholestérolémie normale et cholestérolémie anormale. Regroupant dans cette seconde classe les patients de la classe «cholestérolémie basse» et «cholestérolémie très élevée», elle sacrifie la notion d'ordre et passe à une échelle nominale.

Pour gagner en simplicité de représentation, Stéphanie accepte de perdre de l'information, d'abord en regroupant les données individuelles en classes, puis à chaque passage d'une échelle à l'autre. Il est bien clair que la démarche inverse ne peut se faire et qu'une personne qui ne dispose que de l'information que fournit une échelle nominale ne peut reconstruire ni les échelles ordinales ni les échelles par intervalle.

2.3 Les distributions de fréquences

Comment est-il possible de présenter la pression artérielle dans un groupe de 100 personnes ou le poids des écoliers d'une école primaire de manière à pouvoir déterminer l'importance relative des valeurs de ces mesures? L'organisation des données sous forme de distribution de fréquences les rend plus claires et plus intelligibles.

Après avoir défini une échelle de classification et établi les différentes classes nécessaires, il faut répartir les observations dans ces classes. Le nombre d'observations regroupées dans une classe est appelé «fréquence absolue» ou «effectif» de cette classe. L'ensemble des classes d'une échelle avec leur fréquence constitue ce que l'on appelle une **distribution de fréquences** présentée sous forme de tableau.

EXEMPLE 2.14

Le service de santé de l'école que fréquente Yako voulait connaître l'âge des parents des enfants de la maternelle et de la première année. On a recueilli l'âge des 78 parents. La distribution de fréquences de l'âge des parents est présentée dans le tableau 2.3, à la page 24.

On peut noter qu'un tableau de distribution de fréquences comporte généralement quatre colonnes. La première présente les classes; la deuxième, l'effectif ou la fréquence absolue de chaque classe; la troisième, la fréquence relative qui présente en pourcentage la répartition des effectifs dans les différentes classes (exemple pour la classe 20-24 ans: $10/78 = 12,8\%$); et la dernière, la fréquence relative cumulée, qui additionne la fréquence relative des classes (exemple pour la classe 25-29 ans: $12,8\% + 14,1\% = 26,9\%$). Les tableaux de fréquences, comme tous les tableaux, doivent être numérotés, porter un titre concis qui précise l'information présentée et fournir la source des données utilisées.

TABLEAU 2.3 ■ La distribution de fréquence

Âge (ans)	Effectif	Fréquence relative (%)	Fréquence relative cumulée (%)
20-24	10	12,8	12,8
25-29	11	14,1	26,9
30-34	24	30,8	57,7
35-39	12	15,4	73,1
40-44	15	19,2	92,3
> 44	6	7,7	100,0
Total	78	100,0	

Source : Service de santé, école Sainte-Cunégonde, octobre 2016.

2.4 Les représentations graphiques des distributions de fréquences

Les représentations graphiques des distributions de fréquences produisent une évocation visuelle et mettent en relief les éléments essentiels d'un tableau en permettant de saisir plus rapidement et plus facilement les grands traits d'une distribution. Les modes de représentation graphique varient selon le type d'échelle.

2.4.1 L'échelle nominale ou ordinale

Dans ce groupe, on distingue principalement trois sortes de représentations graphiques.

Le diagramme en secteurs proportionnels

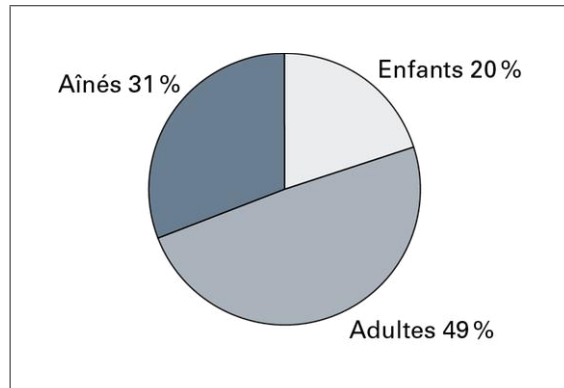
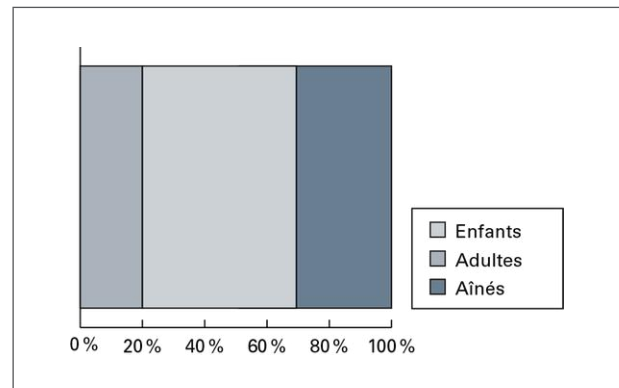
Aussi appelés « tartes », les diagrammes en secteurs proportionnels représentent les proportions des valeurs prises par une variable. Ils ne permettent cependant pas de représenter aisément plusieurs séries de données.

EXEMPLE 2.15

Lisa a constaté qu'en 2016, 20 % de ses consultations ont concerné des enfants, 49 %, des adultes, et 31 %, des aînés. Le diagramme en secteurs proportionnels de la figure 2.1 illustre cette distribution de fréquences.

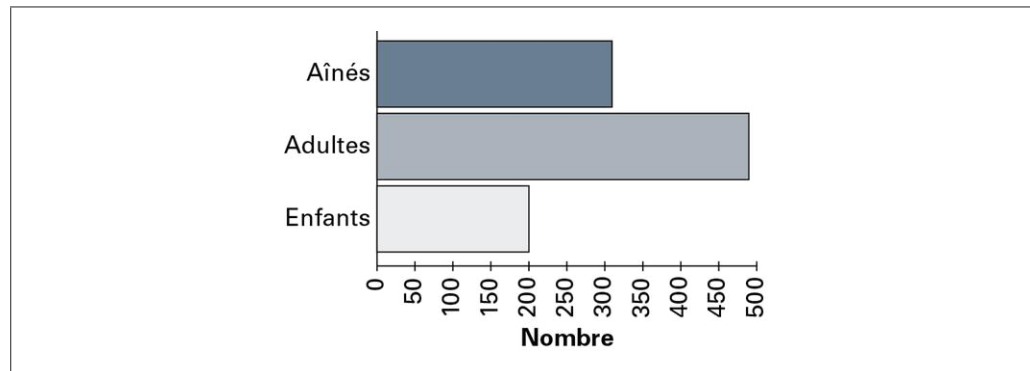
Le diagramme en barres proportionnelles

Cette représentation graphique est analogue à la représentation par tarte, sauf que les fréquences sont partagées sur une barre (*voir la figure 2.2*).

FIGURE 2.1 ■ Un diagramme en secteurs proportionnels**FIGURE 2.2 ■ Un diagramme en barres proportionnelles**

Le diagramme en barres

Dans ce type de présentation graphique, les données représentées sous forme de barres sont disposées horizontalement ou verticalement. Chaque barre représente la fréquence de la classe d'une variable (voir la figure 2.3).

FIGURE 2.3 ■ Un diagramme en barres

2.4.2 L'échelle par intervalle

Pour l'échelle par intervalle, les histogrammes et les polygones de fréquences sont les plus utilisés.

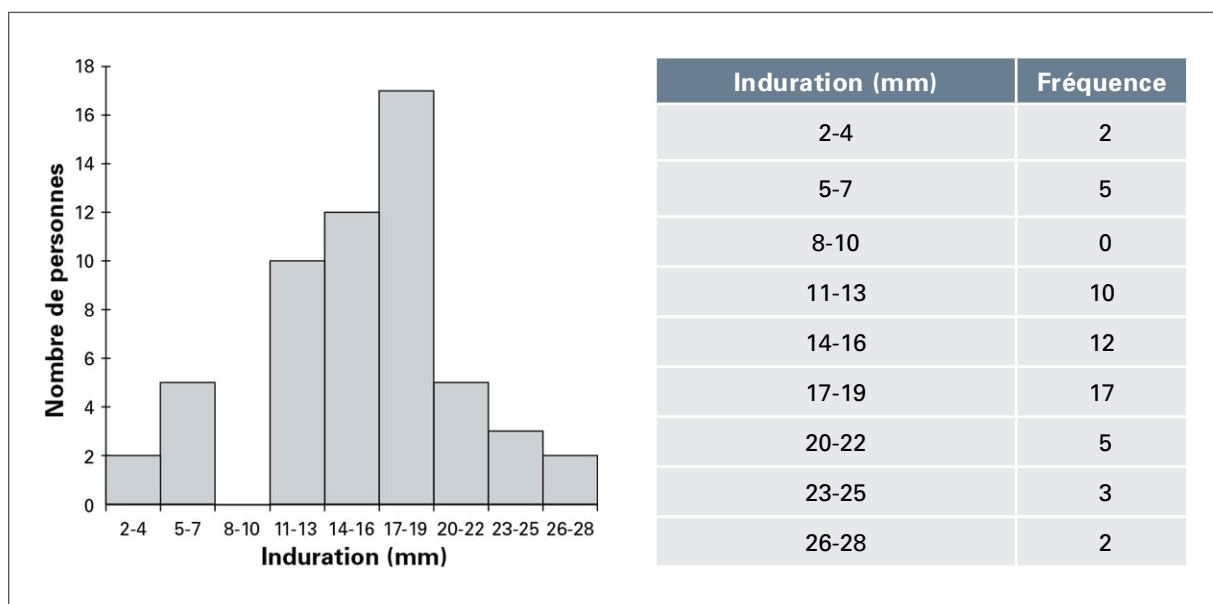
L'histogramme

Cette forme de représentation graphique est utilisée pour les variables quantitatives continues. L'histogramme se construit dans un système d'axes rectangulaires. Pour chaque rectangle, les conditions suivantes doivent être respectées.

- Les rectangles se suivent dans l'ordre des classes (par exemple, pour la variable *âge*: 0-1 an, 2-3 ans, 4-5 ans, 6-7 ans).
- Chacune des bases des rectangles coïncide avec l'intervalle de la classe correspondante.
- Chacune des aires des rectangles (base \times hauteur) mesure la fréquence de la classe correspondante. En pratique, la plupart des auteurs prendront soin de diviser les données en classes égales et, donc, d'accorder une base égale à chaque rectangle. Ainsi, on pourra se fier à la hauteur des rectangles pour évaluer la fréquence de chaque classe.

Dans la figure 2.4, une distribution de fréquences est représentée au moyen d'un histogramme et sous forme de tableau.

FIGURE 2.4 ■ Un histogramme et une distribution de fréquences correspondante



Source: Archives de la polyclinique Milo, 2010.

Le polygone de fréquences

Ce mode de présentation graphique peut être utilisé pour représenter les variables continues ou les variables discrètes qui ont été regroupées. On le réserve cependant aux variables continues en raison de l'impression de continuité qu'il procure. On obtient un polygone de fréquences en passant une ligne par le milieu des sommets des rectangles constituant un histogramme (*voir la figure 2.5*).

Le graphe en lignes

On peut construire ce graphe à partir d'une variable y en fonction d'une variable x .

EXEMPLE 2.16

L'évolution du nombre des consultations de Lisa au cours des trois dernières années peut se lire dans le graphe en lignes présenté à la figure 2.6.

FIGURE 2.5 ■ Un polygone de fréquences: test de dépistage de la tuberculose

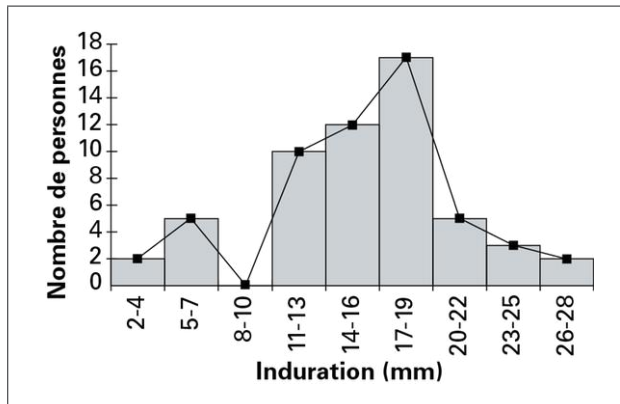
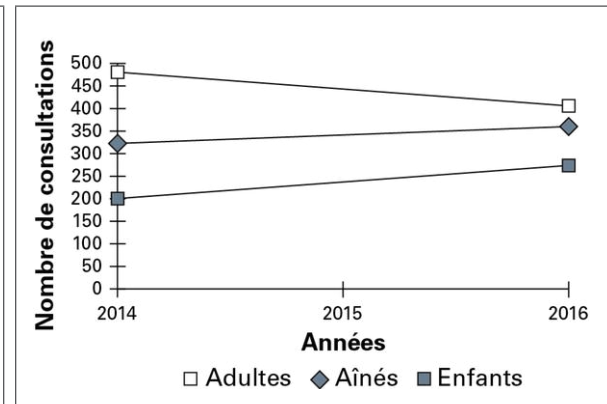


FIGURE 2.6 ■ Un graphe en lignes: évolution des consultations



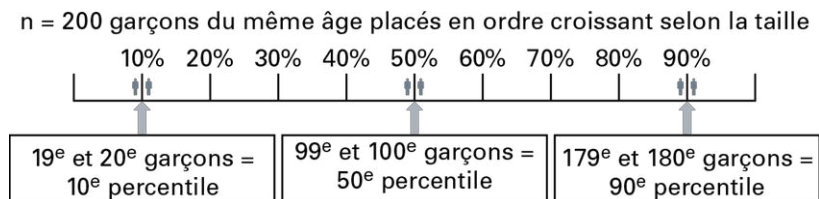
Le graphe de percentiles

Les percentiles partagent une distribution en 100 parties égales entre elles.

EXEMPLE 2.17

Danielle constitue un groupe de 200 garçons de l'école de Yako, tous du même âge, et les ordonne selon leur taille. En commençant par le plus petit, les 19^e et 20^e garçons ont une taille qui correspond au 10^e percentile pour ce groupe. De part et d'autre de ces garçons, on trouve 18 et 180 garçons, soit respectivement 9% et 90% du groupe. De la même façon, si Danielle considère les 179^e et 180^e garçons dans le même ordre, leur taille représentera le 90^e percentile.

Cette figure illustre la façon dont Danielle a réparti en percentiles la taille des 200 garçons du même âge.



C'est selon cette approche que sont construites pour chaque âge les tables de percentiles comme celle représentée à la figure 2.7. L'âge y est indiqué en abscisse,

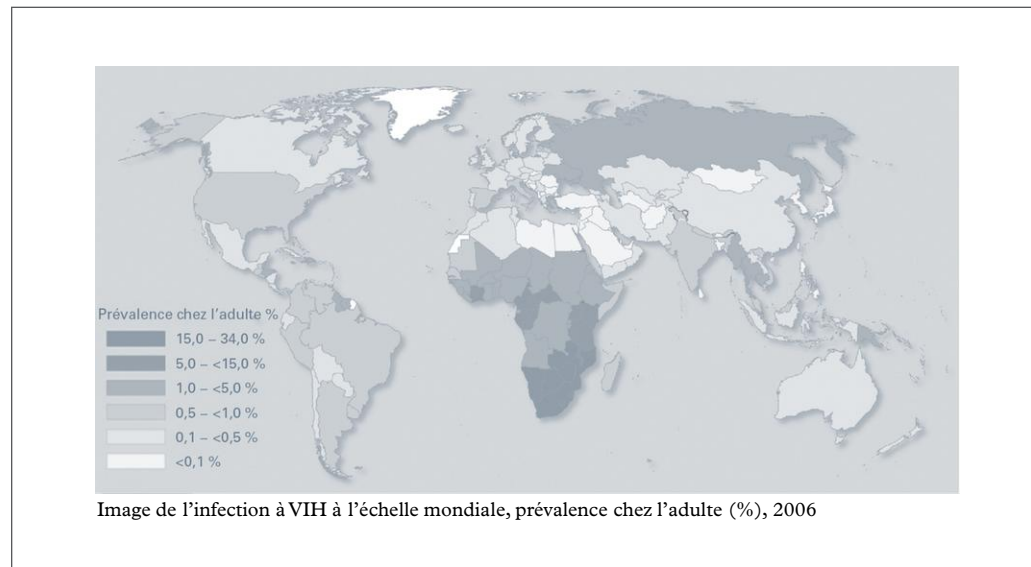
2.4.3 Les autres représentations graphiques

Ce bref relevé des représentations graphiques des distributions de fréquences n'est pas exhaustif. On se limite ici à la forme la plus simple de ces modèles de représentations. La littérature scientifique fait régulièrement appel à une utilisation plus compliquée de celles-ci, par exemple la superposition d'un graphe en ligne sur un histogramme pour illustrer la relation entre les distributions de fréquences de deux variables. La compréhension des représentations graphiques présentées ici vous permettra cependant de saisir ces formes plus élaborées. Il s'avère aussi parfois nécessaire et utile de représenter des données d'une autre façon.

La représentation cartographique

La représentation cartographique est très utile pour décrire un phénomène en fonction du lieu (province, département, village, secteur, etc.). La figure 2.8 offre un exemple de représentation cartographique de l'infection par le virus de l'immunodéficience humaine à l'échelle mondiale, prévalence chez l'adulte (%), en 2006.

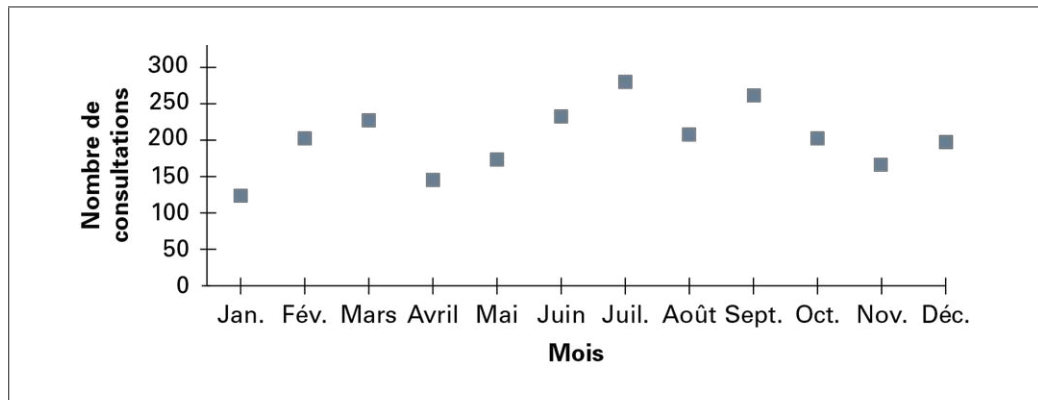
FIGURE 2.8 ■ Une représentation cartographique



Source : ONUSIDA, Organisation mondiale de la Santé, *Unir le monde contre le sida*, 2006.

Le graphe en points

Ce graphe est surtout utilisé lorsque le but de l'étude ne consiste pas à faire des extrapolations à partir des valeurs trouvées. Chaque point est caractérisé par son abscisse et son ordonnée sur des axes de repère donnés ou sur une représentation cartographique. La figure 2.9 constitue un exemple de ce type de représentation.

FIGURE 2.9 ■ Un graphe en points : consultations en 2015

2.4.4 Les règles générales de représentation graphique

Les graphiques doivent être aussi simples que possible. Chacun doit comporter un titre clair, concis et précis. Les échelles et les unités doivent être spécifiées ; les sources des données, indiquées ; et une légende est ajoutée si elle se révèle nécessaire à la compréhension du graphique. Il faut cependant éviter toute surcharge.

2.5 Les mesures descriptives générales d'un ensemble de données

Les mesures descriptives générales d'un ensemble de données permettent de produire ou de résumer le même type d'information utile et intelligible que celle obtenue à l'aide d'une distribution de fréquences et de sa représentation graphique.

Stéphanie effectue de la surveillance nutritionnelle dans l'école que fréquente Yako. Au début de l'année, elle pèse tous les enfants. L'école compte 20 groupes de 15 élèves chacun. Comment peut-elle décrire, de façon générale, le poids des enfants de chacun des groupes ? Les mesures de tendance centrale et les mesures de dispersion sont utilisées à cette fin.

2.5.1 Les mesures de tendance centrale

Une valeur centrale se caractérise par le fait que toutes les valeurs observées tendent à se rassembler autour d'elle. On peut imaginer une valeur centrale comme une sorte de valeur typique autour de laquelle gravitent les valeurs observées d'une variable. On distingue trois sortes de **mesures de tendance centrale** : la moyenne, la médiane et le mode.

La moyenne

Il existe plusieurs types de moyennes. Les principaux types qui figurent dans la littérature médicale et psychosociale sont les moyennes arithmétique, pondérée et géométrique.

La moyenne arithmétique La **moyenne arithmétique** représente la somme des valeurs observées, divisée par le nombre de ces valeurs. Elle s'exprime mathématiquement de la façon suivante :

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

μ = la moyenne arithmétique
 Σ = la sommation des valeurs observées
 x = chaque observation
 $i = 1, 2, \dots, n$
 n = le nombre d'observations

Voici le nombre d'enfants par famille des 15 enfants que compte la classe de Yako :

3, 2, 2, 4, 1, 2, 3, 1, 2, 4, 5, 3, 1, 2, 3

Le calcul de la moyenne arithmétique s'effectue comme suit :

$$\mu = \frac{3+2+2+4+1+2+3+1+2+4+5+3+1+2+3}{15} = \frac{38}{15} = 2,5.$$

La moyenne arithmétique est donc de 2,5 enfants par famille. Notez que le nombre d'enfants par famille est une variable discrète. Une variable discrète présentée en nombre absolu ne peut être traduite par une fraction. Cependant, elle peut s'exprimer à l'aide d'une fraction lorsqu'on en calcule la moyenne.

La moyenne arithmétique présente cependant une lacune. Elle est influencée par les valeurs extrêmes, surtout élevées. Une valeur extrême est une valeur exceptionnellement plus élevée ou moins élevée que la majorité des valeurs observées pour une variable.

EXEMPLE 2.18

Les nombres 2, 4, 6, 8 et 56 ont 11,2 pour moyenne. La valeur 56 influence la moyenne arithmétique de ces valeurs, de sorte que la moyenne 11,2 n'exprime pas réellement la tendance centrale ou le point milieu de la grande majorité des valeurs de cet ensemble de données.

Il est à noter que l'influence des valeurs extrêmes se trouve atténuée si le nombre d'observations est grand.

La moyenne pondérée Lorsque les données sont regroupées en classes, comment peut-on calculer une moyenne? On peut additionner les moyennes des classes, puis les diviser par le nombre de classes, mais l'on introduit ainsi une erreur importante si une classe contient plus d'observations qu'une autre. Pour corriger ce problème, on donne un poids (on pondère) à chaque classe selon son importance (son poids relatif) dans l'ensemble.

EXEMPLE 2.19

Voici le poids des élèves de la classe de Yako; les données sont regroupées en trois classes. Quatre enfants pèsent de 20 à 24 kg; 8 enfants, de 25 à 29 kg; et 3 enfants, de 30 à 34 kg. Pour calculer la moyenne du groupe, il faut faire une supposition quant au poids moyen dans chaque classe. En supposant que le poids est distribué uniformément dans chacune d'elles, on calcule donc la moyenne arithmétique du poids de chaque classe de la façon suivante :

$$\text{Classe 20 - 24 kg: } \frac{20 + 22 + 22 + 22}{4} = 21,5 \text{ kg}$$

$$\text{Classe 25 - 29 kg: } \frac{25 + 25 + 26 + 26 + 27 + 28 + 29 + 29}{8} = 26,9 \text{ kg}$$

$$\text{Classe 30 - 34 kg: } \frac{30 + 31 + 33}{3} = 31,3 \text{ kg}$$

La **moyenne pondérée** se calcule en faisant la somme des moyennes des trois classes multipliée par le nombre d'observations dans chaque classe, divisée par le nombre total d'élèves.

$$\text{Moyenne pondérée: } \frac{(4 \times 21,5) + (8 \times 26,9) + (3 \times 31,3)}{15} = 26,3 \text{ kg}$$

Une autre façon de calculer la moyenne pondérée pour des données groupées en classe serait d'utiliser le point milieu de chaque classe. Cette approche suppose que la répartition selon le poids est symétrique et uniforme à l'intérieur de chaque groupe d'élèves. Le tableau 2.4 présente cette approche.

TABLEAU 2.4 ■ Une illustration de l'approche de la moyenne pondérée

Classes	Point milieu de la classe	Fréquence	Point milieu multiplié par la fréquence
20 - 24 kg	$(20 + 24,9)/2 = 22,5$	4	$22,5 \times 4 = 89,8$
25 - 29 kg	$(25 + 29,9)/2 = 27,5$	8	$27,5 \times 8 = 219,6$
30 - 34 kg	$(30 + 34,9)/2 = 32,5$	3	$32,5 \times 3 = 97,5$
Total		15	406,9

La moyenne s'obtient alors en divisant la somme estimée des valeurs observées par le nombre d'élèves, soit: $406,9/15 = 27,1 \text{ kg}$.

La moyenne géométrique

Cette moyenne, qui utilise le produit des valeurs observées, n'est définie que pour des valeurs positives. Elle permet de réduire l'influence des valeurs extrêmes, surtout basses. La **moyenne géométrique** est particulièrement utilisée dans le contexte des analyses de laboratoire où les distributions sont souvent asymétriques.

La formule générale de calcul de la moyenne géométrique est la suivante :

$$\mu_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

μ_g = la moyenne géométrique
 Π = le produit des valeurs observées
 x_i = chaque observation
 $i = 1, 2, \dots, n$
 n = le nombre d'observations

Ou : $\mu_g = (x_1 \times x_2 \times \dots \times x_n)^{1/n}$

EXEMPLE 2.20

Étant donné le poids des enfants de la classe de Yako, le calcul de la moyenne géométrique s'effectue comme suit :

$$\mu_g = \sqrt[15]{(20)(22)(22)(22)(25)(25)(26)(26)(27)(28)(29)(29)(30)(31)(33)}$$

ou

$$(20 \times 22 \times 22 \times 22 \times 25 \times 25 \times 26 \times 26 \times 27 \times 28 \times 29 \times 29 \times 30 \times 31 \times 33)^{1/15}$$

La moyenne géométrique est donc de 26,1 kg.

Le comportement des moyennes arithmétique et géométrique en présence de valeurs extrêmes Comme le montre le tableau 2.5, le comportement des moyennes arithmétique et géométrique varie quand on introduit des valeurs extrêmes dans un ensemble de valeurs.

TABLEAU 2.5 ■ Des exemples de moyennes en présence de valeurs extrêmes

Série de valeurs	Moyenne arithmétique	Moyenne géométrique
3, 4, 5	4	3,9
3, 4, 5, 16	7	5,6
14, 15, 16	15	15,0
3, 14, 15, 16	12	10,2

La médiane

Cette mesure divise en deux parties égales en nombre l'ensemble des valeurs observées, préalablement mises en ordre croissant. Une moitié des valeurs observées est inférieure à la médiane, et l'autre moitié est supérieure à celle-ci.

La **médiane** dépend du rang des observations disposées en ordre numérique. Elle prend la valeur de l'observation de rang $(n + 1)/2$, où n représente le nombre d'observations.

Contrairement à la moyenne, la médiane n'est pas influencée par les valeurs extrêmes. On la préfère aux autres mesures de tendance centrale lorsque la distribution des valeurs est fortement asymétrique.

Lorsque la valeur n est impaire, il devient facile de choisir la valeur médiane, puisqu'elle est représentée par un chiffre entier.

EXEMPLE 2.21

Les valeurs observées en ordre numérique croissant sont les suivantes :

2, 7, 8, 10, 12, 16, 18

Le calcul du rang s'effectue comme suit :

$$\frac{n+1}{2} = \frac{7+1}{2} = 4$$

La quatrième valeur représente la médiane de cet ensemble de valeurs. La médiane est donc 10.

EXEMPLE 2.22

Les valeurs du diamètre d'induration (mm) mesuré lors d'un test de dépistage de la tuberculose sont les suivantes, en ordre numérique croissant :

2, 4, 5, 5, 5, 8, 8, 8, 9, 10, 11, 11, 12, 13,
13, 13, 13, 13, 14, 15, 15, 15, 16, 17, 17,
17, 17, 17, 18, 19, 19, 19, 20, 20, 20, 21,
22, 23, 23, 23, 23, 24, 25, 26, 28

La médiane = $(45 + 1)/2 = 23^{\text{e}}$ valeur = 16

Lorsque la valeur n est paire, la médiane se situe entre deux valeurs observées et doit être calculée.

EXEMPLE 2.23

Les valeurs observées en ordre numérique croissant sont les suivantes :

24, 28, 32, 40, 50, 60

Le calcul du rang s'effectue comme suit :

$$\frac{n+1}{2} = \frac{6+1}{2} = 3,5$$

La médiane est la valeur comprise entre 32 et 40, soit la valeur qui se situe entre la troisième et la quatrième place.

La médiane = $(32 + 40)/2 = 36$

Le mode

Le **mode** est la valeur qui revient le plus souvent dans un ensemble de valeurs observées.

EXEMPLE 2.24

Les valeurs observées sont les suivantes: 13, 14, 14, 16, 16, 16, 16, 17, 18, 19.
Le mode est donc 16.

Le mode est influencé par les fréquences des observations. Il est plus sensible aux changements et moins stable que la moyenne, surtout si l'on dispose de peu d'observations. Il est possible de le déterminer pour des variables tant qualitatives que quantitatives.

Dans l'exemple 2.24, il n'y a qu'un seul mode; on parle alors d'une distribution unimodale. Il est possible de trouver plusieurs modes dans une même distribution de fréquences. L'exemple 2.25 présente deux modes, c'est-à-dire deux valeurs qui reviennent le plus souvent; il s'agit alors d'une distribution bimodale. Il est possible aussi qu'une distribution n'ait aucun mode.

EXEMPLE 2.25

13, 14, 14, 16, 16, 16, 16, 17, 18, 19, 19, 19, 19

Les modes 16 et 19 reviennent chacun quatre fois. Il s'agit des deux valeurs qui reviennent le plus souvent.

La relation entre les échelles et les mesures de tendance centrale

On a vu que les échelles utilisées pour classer les valeurs des variables sont porteuses d'une plus ou moins grande quantité d'information. Le tableau 2.6 passe en revue les types de mesures de tendance centrale utilisées en fonction des échelles de classification dont on dispose.

TABLEAU 2.6 ■ La relation entre les échelles de classification et les mesures de tendance centrale

Échelle utilisée	Mesures de tendance centrale
Nominale	Mode
Ordinale	Mode, médiane
Par intervalle	Mode, médiane, moyenne

2.5.2 Les mesures de dispersion

Stéphanie a décrit, à l'aide d'une mesure de tendance centrale, le poids des élèves de l'école Sainte-Cunégonde qui, évidemment, n'ont pas tous le même poids. Comment peut-elle réussir à représenter la variation de poids sans énumérer celui de tous les élèves?

Les mesures de tendance centrale ne suffisent pas à caractériser complètement une distribution de fréquences. Voilà pourquoi il est essentiel de définir des **mesures de dispersion** qui fournissent des renseignements sur la variabilité des observations.

EXEMPLE 2.26

Voici la taille, en centimètres (cm), de deux groupes de 12 parents d'enfants fréquentant l'école de Yako.

Groupe A: 155, 160, 164, 165, 170, 170, 170, 170, 175, 176, 180, 185

- La moyenne est de 170 cm.
- Le mode est de 170 cm.
- La médiane est de 170 cm.

Groupe B: 145, 150, 160, 165, 170, 170, 170, 170, 178, 182, 185, 195

- La moyenne est de 170 cm.
- Le mode est de 170 cm.
- La médiane est de 170 cm.

Comme on le constate, les mesures de tendance centrale s'avèrent identiques pour chacun des deux groupes A et B. Pourtant, ces groupes sont différents : les mesures de la taille observées dans ces deux groupes ne sont pas les mêmes. On trouve dans le groupe B des individus de plus petite taille et d'autres de plus grande taille que ceux observés dans le groupe A. Les mesures de dispersion apportent une information descriptive supplémentaire sur la répartition des personnes selon leur taille qui permettra de les distinguer. Les principales mesures de dispersion sont l'étendue, les quantiles, la variance, l'écart type et le coefficient de variation.

L'étendue

L'**étendue** est la différence entre la valeur la plus grande (maximale) et la valeur la plus petite (minimale) d'un ensemble d'observations.

EXEMPLE 2.27 (voir l'exemple 2.26)

- L'étendue du groupe A = $185 - 155 = 30$ cm.
- L'étendue du groupe B = $195 - 145 = 50$ cm.

Cette mesure néglige cependant une partie considérable de l'information : elle ne subit l'influence que des valeurs extrêmes.

Les quantiles

Alors que la médiane donne la valeur centrale d'un ensemble de données, les **quantiles** permettent de diviser l'ensemble en un certain nombre de parties égales. Comme pour la médiane, le calcul des quantiles exige que les données soient ordonnées. Lorsqu'on divise le groupe en quatre parties égales, on obtient des quartiles. S'il est divisé en cinq parties égales, on obtient des quintiles, etc. Si l'on divise l'ensemble en 100 parties égales, on détermine les percentiles.

EXEMPLE 2.28 (voir l'exemple 2.26)

Groupe A : 155, 160, 164, | 165, 170, 170, | 170, 170, 175, | 176, 180, 185

On dira que les parents dont la taille est inférieure à 165 cm appartiennent au premier quartile et que ceux dont la taille est égale ou supérieure à 176 cm appartiennent au quatrième quartile.

Groupe B : 145, 150, 160, | 165, 170, 170, | 170, 170, 178, | 182, 185, 195

On dira que les parents dont la taille est inférieure à 165 cm appartiennent au premier quartile et que ceux dont la taille est égale ou supérieure à 182 cm appartiennent au quatrième quartile.

Pour la surveillance de la croissance des enfants, les courbes du poids ou de la taille tracées en fonction de l'âge sont comparées aux percentiles d'une population de référence.

La variance

Dans un ensemble de données, la valeur de chacune des données d'une variable est plus ou moins proche de sa moyenne. L'idée d'utiliser la distance (ou l'écart) de chaque valeur par rapport à cette moyenne constitue une mesure de dispersion connue sous le nom de **variance**.

EXEMPLE 2.29

Soit le groupe A : 165, 170, 175, 185, 170, 176, 155, 170, 164, 170, 160, 180 cm

La moyenne est de 170 cm. L'écart entre chaque valeur et la moyenne s'obtient en soustrayant la moyenne de chacune des valeurs, par exemple :

$$165 - 170 = -5 \text{ cm};$$

$$170 - 170 = 0 \text{ cm};$$

$$175 - 170 = 5 \text{ cm}.$$

On trouve donc : -5, 0, 5, 15, 0, 6, -15, 0, -6, 0, -10, 10 cm.

Pour exprimer cette variation par un seul chiffre, on pourrait faire la moyenne de ces écarts. Toutefois, puisque le nombre d'écarts positifs est égal au nombre d'écarts

négatifs, on ne peut directement calculer cette moyenne : en effet, la somme des écarts est toujours égale à 0. Pour rendre tous les écarts positifs, on prendra le carré des écarts.

EXEMPLE 2.30

Soit : $-5^2 = 25$;
 $0^2 = 0$; etc.

On trouve donc :

25, 0, 25, 225, 0, 36, 225, 0, 36, 0, 100, 100 cm^2 .

Pour cette série de valeurs, la variance s'obtient de la façon suivante :

$$\sigma^2 = \frac{25+0+25+225+0+36+225+0+36+0+100+100}{12} = 64,3 \text{ cm}^2$$

La formule générale de calcul de la variance est la suivante :

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

σ^2 = la variance
 Σ = la sommation
 x_i = chaque observation
 μ = la moyenne
 n = le nombre d'observations

La variance est donc la somme des carrés des écarts par rapport à la moyenne divisée par le nombre d'observations.

L'écart type

L'**écart type** est la racine carrée de la variance, soit :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

EXEMPLE 2.31

Puisque la variance du groupe A était de 64,3 cm^2 , son écart type est de 8 cm. La variance du groupe B étant de 184 cm^2 , son écart type est de 13,6 cm.

Ces résultats indiquent que la taille des individus est plus homogène dans le groupe A que dans le groupe B.

Lorsqu'on calcule l'écart type dans un sous-ensemble d'une population, soit un échantillon (voir la section 2.6), on utilise la formule suivante :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

C'est qu'on a constaté que, dans un échantillon, l'écart type était une estimation biaisée de l'écart type de la population s'il était calculé de la même façon. On corrige ce biais en utilisant le dénominateur $n - 1$ plutôt que le dénominateur n .

À retenir

Selon que l'objet d'étude est une population ou un échantillon, on utilise une notation différente.

S'il s'agit d'une population, on désigne :

la moyenne par μ ; la variance par σ^2 ; l'écart type par σ .

S'il s'agit d'un échantillon, on désigne :

la moyenne par m ou \bar{x} ; la variance par s^2 ; l'écart type par s .

Le coefficient de variation

Le **coefficient de variation** permet de comparer deux distributions d'une même variable ou de variables différentes. Il s'agit simplement du rapport de l'écart type à la moyenne, lequel s'exprime de la façon suivante :

$$CV = \frac{\sigma}{\mu}$$

EXEMPLE 2.32

Le coefficient de variation du groupe A est : $CV = \frac{8,021 \text{ cm}}{170 \text{ cm}} = 0,0472$ ou 4,72%

Le coefficient de variation du groupe B est : $CV = \frac{13,57 \text{ cm}}{170 \text{ cm}} = 0,0798$ ou 7,98%

On notera que le résultat du calcul du coefficient de variation est un rapport exprimé par une fraction ou un pourcentage. Il permet de comparer la variabilité des ensembles de données dont les unités diffèrent. Grâce au coefficient de variation, on peut par exemple comparer la variation des valeurs de la variable *âge*, qui va de 1 à 100 ans, à la variation de la variable *cholestérol*, dont les valeurs oscillent de 2 à 15 mmol/l. Plus la valeur du coefficient de variation est élevée, plus grande est la variabilité des données observées.

2.6 L'échantillonnage

Pour situer Yako dans le groupe des 10% plus petits quant au poids et à la taille, Danielle a sélectionné un groupe de 200 garçons parmi tous les garçons du même âge et de la même ville que Yako, ce qui lui a permis d'établir des valeurs de référence. Elle aurait pu étudier un groupe d'individus plus nombreux, par exemple tous les garçons du même âge et de la même région que Yako. Cela aurait sans doute augmenté

le degré de certitude de Danielle lorsqu'elle a classé Yako selon son poids et sa taille. Elle a cependant considéré que le groupe des 200 garçons était suffisant pour lui fournir une certitude raisonnable.

Dans le domaine scientifique, il n'existe pas d'absolu; toutes les connaissances sont plus ou moins certaines, et le savoir est fondé sur la probabilité plus ou moins grande d'observer un phénomène. En effet, à partir d'un certain moment, les coûts en temps et en ressources deviennent trop élevés pour justifier que l'on cherche à augmenter encore davantage le degré de certitude. À partir d'un certain seuil, on considère que l'on en sait assez pour agir puisque, de toute façon, une certitude plus grande ne changerait rien aux décisions ni aux mesures à prendre. Une certitude raisonnable suffit. La pratique qui consiste à mener des études (ou des sondages) en utilisant des échantillons découle de cet état de fait.

Un **échantillon** est simplement un sous-ensemble de la population que l'on souhaite étudier. Il faut choisir ce sous-groupe de telle sorte que l'on puisse obtenir une certitude raisonnable qui permet de tirer des conclusions et d'agir en fonction du facteur étudié.

2.6.1 Le choix d'un échantillon

Pour arriver à choisir son échantillon, Danielle est partie des connaissances qu'elle avait du phénomène à étudier dans le but d'établir des valeurs normales (ou valeurs de référence) qui lui permettraient de situer Yako selon son poids et sa taille. Elle a dû fixer certains critères pour guider le choix de l'échantillon afin de s'assurer d'atteindre son objectif. La sélection d'individus du même âge que Yako constitue l'un de ces critères; il s'agit d'un **critère d'inclusion**. Cependant, comme elle s'intéressait aux valeurs normales, elle a décidé de ne pas retenir tout garçon qui présenterait une caractéristique particulière susceptible de venir fausser le résultat. Ainsi, elle a décidé de ne pas inclure dans son échantillon un garçon ayant souffert d'une maladie gastro-intestinale grave à sa naissance qui aurait affecté sa croissance pondérale; il s'agit d'un **critère d'exclusion**.

Tout en tenant compte des critères d'inclusion et d'exclusion, Danielle a également dû déterminer le nombre de personnes à recruter pour obtenir le seuil de certitude qui rendrait son étude utile, soit la plus petite différence qu'elle veut détecter. De ses cours de statistique, elle avait retenu le principe suivant: un grand échantillon permet de détecter une information plus précise qu'un petit échantillon.

Par conséquent, plus on veut augmenter la précision des renseignements recueillis, plus la taille de l'échantillon doit être grande. Par ailleurs, plus il y a de variabilité dans l'information recherchée (par exemple, plus la taille des garçons du même âge que Yako varie d'un individu à l'autre), plus l'échantillon doit être grand. Un grand échantillon permet des analyses statistiques plus puissantes; plus la puissance est grande, plus le résultat obtenu inspirera confiance. Ainsi, ces notions de précision et de **puissance statistique** permettent de quantifier la confiance que l'on peut accorder à un résultat observé dans un échantillon lorsqu'on veut le généraliser à une population. Elles seront approfondies dans les chapitres 10 et 11.

À retenir

La taille de l'échantillon dépendra donc de :

- la plus petite différence que l'on veut détecter;
- la précision souhaitée;
- la puissance souhaitée;
- la variabilité des données.

Il existe des règles de calcul de la taille des échantillons qui permettent de mesurer les différences avec la précision et la puissance voulues en tenant compte de la variabilité des données.

2.6.2 Les méthodes d'échantillonnage

Après avoir déterminé les critères d'inclusion et d'exclusion des personnes à recruter pour constituer l'échantillon et effectué tous les calculs nécessaires, Danielle en arrive à la conclusion qu'elle doit choisir 200 personnes parmi un nombre beaucoup plus grand de garçons du même âge que Yako. Comment s'y prendra-t-elle ?

La solution la plus avantageuse consiste à donner une chance égale à toutes les personnes d'être sélectionnées et à s'en remettre au hasard. Ainsi, il sera probable que les caractéristiques des personnes choisies ne soient pas vraiment différentes de celles des personnes non sélectionnées. On nomme ce procédé **échantillonnage aléatoire simple**.

Par contre, selon le contexte de l'étude, il est possible d'avoir recours à d'autres stratégies d'échantillonnage. Par exemple, dans le cadre d'une étude portant sur l'acné chez les adolescentes, on veut comparer l'effet du chiogène hispide contenu dans le « thé » sur la peau blanche et sur la peau noire, et l'on désire recruter 200 sujets pour réaliser l'étude. Or, parmi les 10 000 candidates possibles, seules 3 000 personnes sont de race noire. Par le procédé de l'échantillonnage aléatoire simple, on arriverait selon toute probabilité à constituer deux groupes composés de 60 adolescentes noires et de 140 adolescentes blanches. Pour résoudre ce problème, il faudrait choisir au hasard 100 jeunes filles parmi les 7 000 candidates blanches et 100 autres parmi les 3 000 candidates noires. On nomme cette façon de procéder **échantillonnage aléatoire stratifié**.

Dans un autre cas où l'on désire mener une étude sur la consommation de suppléments vitaminés auprès des clients d'une compagnie d'assurance, un taux de participation de 10% serait suffisant pour tirer des enseignements concluants. On peut utiliser un outil générateur de chiffres aléatoires pour générer un chiffre de 1 à 10. À défaut d'un tel outil, 10 blocs de bois, numérotés de 1 à 10, sont déposés dans un sac. Une personne tire un bloc au hasard, soit celui portant le numéro 7. Sur la liste

alphabétique des clients de la compagnie, le 7^e nom a été souligné. On a ensuite souligné le 17^e nom, puis le 27^e nom, et ainsi de suite jusqu'à la fin de la liste. Ces clients ont finalement constitué l'échantillon. On nomme cette façon de faire **échantillonnage systématique**.

EXEMPLE 2.33

Emmanuelle, coopérante dans un pays en voie de développement, est chargée d'évaluer le degré d'efficacité de la vaccination contre la rougeole chez les enfants âgés de deux à cinq ans. Il n'existe pas de recensement précis des habitants de la région où elle travaille, et les gens y vivent dans des villages dispersés, sans adresses ni noms de rues.

Elle explique à ses enquêteurs que, dans quelques villages choisis au hasard, ils devront recueillir des informations auprès de 10 enfants par village. Ils se rendront au centre du village, se dirigeront vers la première maison au nord et obtiendront des informations sur tous les enfants âgés de deux à cinq ans habitant cette maison. S'il y en a moins de 10, ils se dirigeront vers la maison voisine située au nord, et ainsi de suite jusqu'à ce qu'ils aient recueilli des informations sur un total de 10 enfants.

On nomme cette façon de procéder **échantillonnage en grappes**.

Il existe d'autres techniques d'échantillonnage. La préoccupation première demeure toujours de recruter le nombre nécessaire de personnes au moindre coût, tout en veillant à ne pas introduire d'erreurs systématiques qui invalideraient les résultats. En d'autres mots, on doit s'assurer que les caractéristiques des individus qui composent l'échantillon observé soient représentatives de l'ensemble des caractéristiques des personnes qui composent la population à qui l'on veut généraliser les résultats. Les méthodes les plus avantageuses sont généralement l'échantillonnage aléatoire simple et l'échantillonnage aléatoire stratifié.

En résumé

Ce chapitre a permis d'exposer les concepts de variable, d'échelle de classification, de tableau de distribution de fréquences et de représentation graphique des données. On y a aussi abordé les paramètres qui permettent de décrire une population, à savoir les mesures de tendance centrale (la moyenne et la médiane en particulier) et les

mesures de dispersion (particulièrement l'étendue, la variance et l'écart type). Finalement, on y a présenté quatre stratégies d'échantillonnage (soit les échantillonnages aléatoire simple, aléatoire stratifié, systématique et en grappes) qui permettent d'obtenir un échantillon représentatif d'une population.

Exercices**Voir la section complète des exercices sur le Web.**

1. Parmi les variables suivantes, lesquelles sont quantitatives continues?
 - a) La religion.
 - b) La durée du mariage.
 - c) Le nombre d'enfants par famille.
 - d) La langue maternelle.
 - e) L'âge.
2. Voici quelques échelles de classification :
 - A. Niveau d'instruction : primaire, secondaire, collégial, universitaire.
 - B. Évolution d'un cancer : stade I, stade II, stade III, stade IV.
 - C. Diagnostic : malaria, schistosomiase, diarrhée, rougeole.
 - D. Nombre de patients admis dans un hôpital tous les jours : < 20 , $20-39$, $40-59$, $60-79$, > 79 .Établissez de quel type d'échelles il s'agit, en choisissant une réponse parmi les suivantes.
 - a) «A» est ordinale, «B» est ordinale, «C» est nominale, «D» est par intervalle.
 - b) «A» est nominale, «B» est ordinale, «C» est nominale, «D» est par intervalle.
 - c) «A» est ordinale, «B» est ordinale, «C» est nominale, «D» est continue.
 - d) «A» est par intervalle, «B» est par intervalle, «C» est nominale, «D» est discrète.
 - e) «A» est nominale, «B» est nominale, «C» est nominale, «D» est par intervalle.
3. On vous indique que dans le groupe 1, l'écart type de l'âge est 28,9 ans et que dans le groupe 2, l'écart type de l'âge est de 33,5 ans. Dans lequel de ces deux groupes l'âge des individus observés varie-t-il le plus?
 - a) Le groupe 1.
 - b) Le groupe 2.
4. Quelle méthode d'échantillonnage offre la meilleure assurance que les individus qui composent l'échantillon seront représentatifs de la population à l'étude?
 - a) L'échantillonnage en grappes.
 - b) L'échantillonnage systématique.
 - c) L'échantillonnage aléatoire.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 3

Les mesures et les modèles d'études en épidémiologie

Collaboration à la 3^e édition :

Helen Trottier

Antoine Lutumba Ntetu

Au terme de ce chapitre, vous pourrez :

- utiliser les outils épidémiologiques qui permettent d'établir les bases du jugement clinique concernant les risques auxquels sont exposés les patients.

Objectifs spécifiques	Contenu
3.1 Utiliser adéquatement les mesures générales de fréquence et de densité.	3.1 Les mesures de fréquence 3.1.1 La proportion 3.1.2 Le taux 3.1.3 Le ratio 3.1.4 L'indice
3.2 Interpréter correctement les mesures de fréquence et de densité en épidémiologie.	3.2 Les mesures en épidémiologie 3.2.1 La prévalence 3.2.2 Le taux d'incidence 3.2.3 L'incidence cumulée 3.2.4 La cote 3.2.5 Une synthèse sur les mesures en épidémiologie 3.2.6 La population statique ou dynamique, ouverte ou fermée
3.3 Expliquer les principaux indicateurs de santé d'une population.	3.3 Des mesures de mortalité et de morbidité 3.3.1 La mortalité 3.3.2 La morbidité 3.3.3 L'espérance de vie
3.4 Expliquer les principales méthodes d'ajustement des mesures de fréquence: l'ajustement direct et l'ajustement indirect.	3.4 L'ajustement des mesures
3.5 Décrire les modèles de base des études épidémiologiques.	3.5 Les modèles (ou types) d'études en épidémiologie 3.5.1 L'étude transversale 3.5.2 L'étude longitudinale 3.5.3 L'étude cas-témoins
3.6 Décrire la classification des études épidémiologiques, leur finalité, leurs avantages et leurs désavantages.	3.6 Une classification des études épidémiologiques 3.6.1 Les études à visée étiologique 3.6.2 Les études descriptives 3.6.3 Une classification des types d'études selon leur finalité



Mise en situation

Mirabelle se présente au service de consultation sans rendez-vous de la polyclinique Milo. Cette jeune fille de 22 ans se plaint de crampes abdominales, de diarrhées fréquentes, de fatigue et d'amaigrissement. Lisa, l'omnipraticienne qui est de garde ce soir-là, la reçoit et entend l'histoire qui suit.

Mirabelle, qui vit actuellement seule, a repris ses études collégiales depuis deux ans; elle les avait interrompues deux ans auparavant pour aller vivre avec son ami Jeremy à New York, dans le Bronx. Elle l'a quitté quelques mois plus tard, après avoir découvert qu'il était bisexuel. Jeremy consommait des drogues intraveineuses. Mirabelle, qui le faisait aussi à l'occasion, a mis fin à sa consommation après s'être séparée de Jeremy. Avant de revenir au Québec, elle a voyagé un an et demi aux États-Unis et au Mexique.

À l'examen physique, Lisa découvre une jeune fille manifestement amaigrie, légèrement déshydratée, avec plusieurs chaînes ganglionnaires hypertrophiées. Elle lui prescrit quelques analyses de laboratoire. Lorsque Mirabelle revient

voir Lisa pour en connaître les résultats, elle s'attend déjà à recevoir un diagnostic d'infection par le virus de l'immunodéficience humaine (VIH) et à apprendre qu'elle est atteinte du syndrome d'immunodéficience acquise (sida).

Pertinence clinique

La démarche clinique ne se résume pas à comparer des signes, des symptômes et des valeurs de laboratoire à une norme. Le thérapeute doit chercher à repérer, dans les antécédents, les habitudes de vie et l'environnement de la personne qui consulte, tout ce qui constitue un risque pour sa santé. Ainsi, devant une femme d'affaires de 62 ans, toujours entre deux réunions et deux avions, il s'intéressera particulièrement à sa condition cardiovasculaire. Par contre, il recherchera plutôt des signes d'épuisement chez une jeune professionnelle toujours coincée entre plusieurs contrats, le soin des enfants et celui du ménage, et dont le conjoint est souvent absent. Le vieillard, veuf, toujours seul à la maison, évoquera une autre série de problèmes de santé possibles.

3.1 Les mesures de fréquence

Albert est un collègue de Lisa. Il travaille dans un centre local de services communautaires (CLSC) en région. Il y a quelques jours, on lui a transmis le communiqué suivant : huit cas d'une grippe particulièrement virulente ont été déclarés chez des personnes âgées du village de Sainte-Cunégonde. Le lendemain, un autre communiqué l'informait qu'il y avait 10 cas de la même grippe dans le village d'Arnoldvénituk, situé à 20 km du premier. Cette information partielle ne permet cependant pas à Albert de saisir l'importance du problème dans chacun des villages.

Des chiffres isolés (8 cas, 10 cas) ont peu de signification lorsqu'on doit mesurer l'importance d'un phénomène: des données supplémentaires sont nécessaires pour que ces chiffres trouvent tout leur sens. Ainsi, sachant que la population de Sainte-Cunégonde et celle d'Arnoldvérituk comptent respectivement 16 et 524 personnes âgées, Albert peut saisir l'ampleur du problème à Sainte-Cunégonde et s'expliquer la très grande inquiétude des familles de ce village.

Quand on fait une comparaison entre le nombre de cas avérés et la population concernée, on établit un rapport. Le **rapport** est une expression générale de la relation entre deux quantités qui peuvent indifféremment appartenir ou non au même ensemble. Un rapport peut se présenter sous la forme d'une proportion, d'un taux, d'un ratio ou d'un indice.

EXEMPLE 3.1

À Sainte-Cunégonde, le rapport entre le nombre de personnes atteintes de la grippe et la population des personnes âgées est de $8/16$ ou $0,5$ (50 %), alors qu'à Arnoldvérituk, ce rapport est de $10/524$ ou $0,019$ (1,9 %).

Ici, les personnes malades font partie intégrante de la population des personnes âgées. On peut aussi établir un rapport entre le nombre de personnes âgées et le nombre de bancs publics dans un village. Dans ce cas-là, les deux entités mises en rapport ne proviennent pas du même ensemble: d'une part, on a l'ensemble des personnes âgées, et de l'autre, l'ensemble des bancs publics.

3.1.1 La proportion

Une **proportion** est un rapport dans lequel les deux entités proviennent du même ensemble et où le numérateur est inclus dans le dénominateur. Le rapport des personnes âgées grippées à l'ensemble des personnes âgées du village est donc une proportion. La valeur d'une proportion se trouve toujours comprise entre 0 et 1, c'est-à-dire entre 0 % et 100 %. Ainsi, si aucune personne âgée n'avait la grippe à Arnoldvérituk, la proportion des personnes âgées grippées serait de $0/524$, soit 0 ou 0 %. Par contre, si toutes les personnes âgées étaient atteintes de la grippe, la proportion serait de $524/524$, soit 1 ou 100 %.

EXEMPLE 3.2

- Des 5 000 personnes qui font partie de la liste des patients d'une clinique, 100 ont des problèmes cardiaques. La proportion de ces dernières est donc de $0,02$ (ou 2 %).
- À New York, pendant que Mirabelle y habitait, on a fait une étude auprès des utilisateurs de drogues intraveineuses (UDI). Sur 324 UDI habitant le Bronx, 130 possédaient des anticorps contre le VIH, ce qui représente une proportion de 40 %.

3.1.2 Le taux

Le **taux** est un rapport qui mesure la vitesse ou la force (intensité) de survenue d'un événement. Le taux est une mesure instantanée de densité. Le numérateur d'un taux,

comme celui d'une proportion, dénombre les événements qui sont survenus : la différence entre les deux réside dans le dénominateur. Le dénominateur de la proportion inclut des personnes ou des entités, tandis que dans le cas du taux, il s'agit du cumul du temps d'observation de chaque personne ou entité analysée qui est à risque de présenter le phénomène étudié. Dans ce cas, le dénominateur sera exprimé en unités de temps (jours, mois, années, etc.). Si les phénomènes observés sont des personnes, on parlera de « personnes-temps » ; si ce sont des distances, on parlera de « distances-temps » (« km-h » ou « m-sec », par exemple).

Lisa s'inquiète du phénomène d'obésité qui devient de plus en plus important chez ses patients. Elle aimerait bien savoir si la consommation de boissons gazeuses est susceptible de faire partie des sources du problème. Pour étudier la question, elle demande à un agent de recherche d'aller passer la journée au restaurant d'à côté et d'observer les clients. Elle lui confie la tâche de calculer le nombre de cannettes de boissons gazeuses consommées par chaque client ainsi que le temps que chacun passe dans le restaurant (le temps pendant lequel il est à risque de consommer). Elle lui demande aussi de classer les clients dans les catégories « obèses » et « non-obèses » selon une règle simple. Elle obtient ainsi pour chaque groupe la vitesse moyenne de consommation de boissons gazeuses, que nous appellerons « taux de consommation ».

Pour obtenir le dénominateur, l'agent de recherche cumule le temps d'observation que chacun des clients passe sous suivi. Dans la journée, 50 personnes obèses sont venues au restaurant : certaines y ont passé moins de 30 minutes et d'autres plus de 2 heures. Le cumul de l'ensemble des temps de suivi a permis d'estimer la durée d'observation totale à 70 heures. Certaines personnes ont bu plus de trois cannettes de boissons gazeuses, alors que d'autres n'en ont pris aucune. Le total des consommations pour le groupe s'est élevé à 40. Chez les 450 personnes non obèses qui ont été observées, on a enregistré 200 consommations de boissons gazeuses, pour une durée d'observation totale de 600 heures.

On est donc en mesure d'estimer le taux de consommation de boissons gazeuses dans les deux groupes :

- Obèses : 40 boissons gazeuses/70 personnes-heures ou
0,57 boisson gazeuse/personne-heure.
- Non-obèses : 200 boissons gazeuses/600 personnes-heures ou
0,33 boisson gazeuse/personne-heure.

Lisa arrive donc à la conclusion que les personnes obèses consomment plus de boissons gazeuses que les personnes non obèses si, bien sûr, son observation est représentative de ce qui se passe dans le reste de la population !

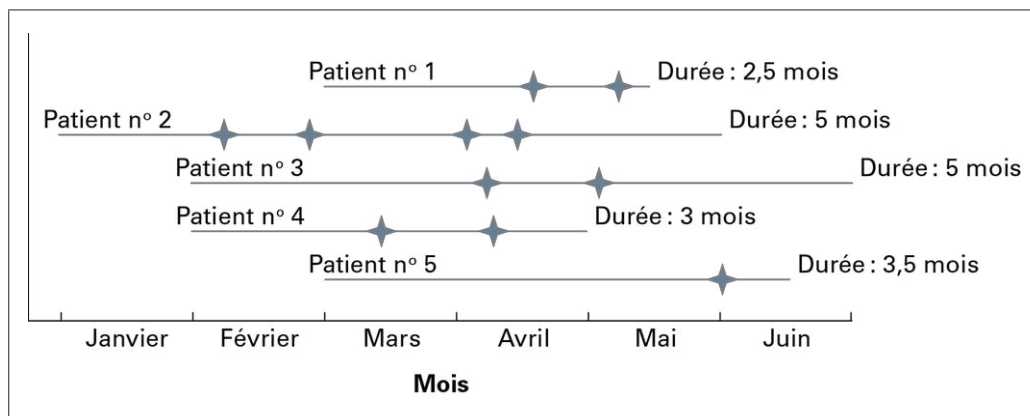
EXEMPLE 3.3

Jean, un clinicien, comptabilise l'apparition des maladies opportunistes chez cinq de ses patients pour qui il a posé un diagnostic de sida, afin d'évaluer le taux d'incidence des infections opportunistes chez ces derniers. Son observation a donné lieu au diagramme de la figure 3.1, dans lequel la longueur de chaque ligne horizontale est proportionnelle à la durée d'observation de chaque personne

EXEMPLE 3.3 (suite)

(la durée dépend du moment où le patient a commencé à participer à l'étude ainsi que de celui où il a cessé d'y participer, soit en raison de son décès ou d'un arrêt de suivi). Les patients n'ont donc pas commencé à prendre part à l'étude au même moment, pas plus qu'ils n'ont cessé d'y participer au même moment. Par ailleurs, Jean arrête de comptabiliser le temps pour les patients qui ont été soignés à l'extérieur durant un certain temps, comme c'est le cas pour le patient numéro cinq. Les étoiles représentent la survenue d'infections. Pour calculer le taux d'incidence, il faut rapporter le nombre total d'événements survenus (infections opportunistes) sur la durée de suivi de l'ensemble des patients étudiés. Au total, chez ses 5 patients, Jean a observé 11 infections opportunistes pendant une durée d'observation de 19 mois (la somme des durées pendant lesquelles chaque patient a participé à l'étude). Cela donne un taux de 11 infections pour 19 personnes-mois, soit $0,58 \text{ mois}^{-1}$.

FIGURE 3.1 ■ L'occurrence des maladies opportunistes chez les personnes ayant le sida



3.1.3 Le ratio

Un **ratio** est un rapport dans lequel le numérateur et le dénominateur proviennent du même ensemble, mais où le numérateur n'est pas compris dans le dénominateur. On note que le ratio peut prendre n'importe quelle valeur et n'a pas d'unité. Par exemple, si, des 324 personnes sur qui portait l'étude menée dans le Bronx, 288 étaient des hommes, le ratio hommes/femmes était donc de $288/36$ ou de $8/1$.

3.1.4 L'indice

Un **indice** est un rapport utilisé lorsque le numérateur n'est pas compris dans le dénominateur et que l'un et l'autre renvoient à deux événements distincts. L'indice s'avère utile lorsque le dénominateur n'est pas bien connu : on utilise alors un dénominateur qui se rapproche de la réalité pour obtenir une mesure.

EXEMPLE 3.4

La fréquence relative des décès maternels dus à une cause puerpérale devrait comporter au dénominateur le nombre total d'accouchements. Ce nombre est généralement mal connu. On peut alors utiliser le nombre de naissances vivantes, qui est généralement disponible. La **mortalité maternelle** s'exprime donc comme suit :

$$\frac{\text{Décès maternels dus à une cause puerpérale}}{\text{Nombre de naissances vivantes}}$$

3.2 Les mesures en épidémiologie

Les proportions, les taux, les ratios et les indices sont très utilisés en épidémiologie. Dans le domaine de la santé, chacune des mesures porte un nom qui décrit précisément son usage.

3.2.1 La prévalence

Les cas prévalents constituent le nombre existant de malades dans une population à un point fixe du temps.

La **prévalence** (P)* d'une maladie est la proportion des personnes affectées par cette maladie à un moment donné dans la population. Elle s'exprime de la manière suivante :

$$P = m/N$$

où m est le nombre de personnes malades, et N , le nombre de personnes dans la population.

La prévalence indique simplement l'ampleur d'un problème à un moment précis. Elle peut être obtenue en menant une enquête ponctuelle puisque c'est une mesure qui correspond à une photographie instantanée d'une situation dans une population. Certains auteurs parleront alors de prévalence instantanée. Comme le mentionne Ancelle¹, cette notion d'instantanéité doit être prise au sens large. La collecte d'information pour mesurer la prévalence s'échelonne nécessairement sur une certaine période de temps qui est généralement négligeable par rapport à la durée de la maladie d'intérêt. Fletcher et ses collaborateurs² feront la distinction entre la prévalence instantanée et la prévalence de période qui se rapporte aux cas présents à n'importe quel moment d'une période de temps spécifique.

* Certains auteurs utilisent également les termes proportion de prévalence, taux de prévalence (le terme taux ne devrait pas être utilisé, car la prévalence n'est pas un taux, mais une proportion) ou prévalence relative pour désigner la proportion des personnes affectées par une maladie à un moment précis.

EXEMPLE 3.5

À l'établissement collégial que fréquente Mirabelle, l'équipe de santé a obtenu le consentement de 300 étudiants choisis au hasard pour le dépistage du VIH. Au sein de cette population, deux personnes étaient séropositives. Le nombre de cas prévalents dans cette population au moment de l'enquête était donc de 2, et la prévalence, de $2/300$ (2 sur 300) ou 0,7%.

3.2.2 Le taux d'incidence

Le **taux d'incidence** (I), aussi nommé densité d'incidence, est une mesure de la vitesse, de la force ou de l'intensité de propagation d'une maladie dans une population. Il s'exprime de la façon suivante :

$$I = n/T$$

où le numérateur (n), à savoir les événements comptabilisés, représente le nombre de nouveaux cas survenus depuis le début de l'étude jusqu'à sa fin. Au dénominateur, on trouve le cumul du temps écoulé entre le début et la fin de l'observation de chaque personne participant à l'étude et susceptible (à risque) de contracter la maladie étudiée (T) : il s'agit en fait du cumul du temps d'attente observé avant la survenue de chaque nouveau cas ou avant la fin de la période d'observation. L'observation prend fin quand la personne tombe malade ou quand l'étude se termine.

EXEMPLE 3.6

Lisa a deux patients qui ont été exposés à des produits radioactifs dans le cadre de leur travail ; elle décide de les suivre pour estimer la densité d'incidence du cancer chez ces derniers. Le premier patient a été exposé aux produits toxiques le 1^{er} juillet 2012 et a été suivi jusqu'au 31 décembre 2016 (durée d'observation de 4,5 années). Il n'a pas été atteint d'un cancer. Par contre, l'autre patient a été exposé aux produits le 30 avril 2013 et a commencé à souffrir d'un cancer le 30 avril 2014 (durée d'observation d'un an). Le taux d'incidence du cancer est alors de 1 personne/(4,5 + 1 personne-année), soit $1/5,5$ année ou $0,18$ année⁻¹. Le nombre que l'on trouve au dénominateur est une somme de temps calculée à partir de la participation de chaque personne à l'étude. On désigne ce nombre sous le nom de « personnes-temps ». Il est à noter qu'ici, on n'étudie pas un phénomène récurrent. Lorsque la personne est atteinte du cancer, elle cesse de contribuer au cumul des personnes-temps puisqu'elle n'est plus à risque (elle a déjà la maladie).

3.2.3 L'incidence cumulée

L'**incidence cumulée** (IC) est une proportion. Il s'agit d'une probabilité ou d'un risque de survenue d'un événement. Elle s'exprime de la façon suivante :

$$IC = n/R$$

où n est le nombre de nouveaux cas apparus pendant la période d'observation, et où R représente le nombre de personnes susceptibles de devenir des cas, en observation au début de l'étude. Cela est possible quand toutes les personnes en observation sont suivies pendant la période complète de l'étude. En pratique, il arrive souvent que l'on perde de vue certaines personnes. Le calcul du risque devient alors un peu plus compliqué³. Même si la notion de temps n'est pas prise en considération dans le calcul de la proportion, il faut préciser le temps de suivi. Par exemple, si l'on prend 12 heures pour dénombrer les cas de gastroentérite à la suite d'un repas de noces, on aura moins de malades que si l'on prend 2 jours. Ainsi, au résultat du calcul de l'incidence cumulée il faut toujours associer le temps de suivi.

EXEMPLE 3.7

Simone, une jeune épidémiologiste, s'est intéressée à l'infection par le virus du papillome humain (VPH) chez les jeunes femmes. Il y a quelques années, elle a recruté 120 étudiantes universitaires qu'elle a suivies pendant 2 ans, et 40 d'entre elles ont contracté une infection par le VPH. Au début de l'étude, les jeunes femmes n'étaient pas infectées (si une personne avait été infectée au départ, elle n'aurait pas été à risque de contracter la maladie). L'incidence cumulée était donc de 40/120 ou de 33 % pour 2 ans. Ainsi, la probabilité de contracter une infection par le VPH a été estimée à 33 % sur 2 ans. Si la durée de suivi avait été de trois ans, l'incidence cumulée aurait été plus élevée, d'où l'importance de mentionner la période de suivi lorsque l'on reporte une incidence cumulée, même si la proportion en soi n'a pas d'unité.

3.2.4 La cote

Si une caractéristique divise une population en deux groupes, par exemple les malades et les non-malades, le ratio malades/non-malades porte le nom de **cote**. De même, si une exposition divise une population en deux groupes, les exposés et les non-exposés, le ratio exposés/non-exposés porte aussi le nom de cote.

EXEMPLE 3.8

Sur 120 étudiantes universitaires, Simone a trouvé que 40 avaient été infectées par le VPH dans le cadre de son étude et que 80 n'avaient pas contracté le virus. La cote de l'infection est donc de 40/80 ou 0,5.

Le tableau 3.1 résume les caractéristiques des mesures utilisées en épidémiologie.

TABLEAU 3.1 ■ Les caractéristiques des mesures utilisées

Mesure	Usage	Type de rapport	Unité	Domaine de variation
Prévalence (proportion de prévalence)	Portrait instantané	Proportion	Aucune	0–1
Taux d'incidence	Vitesse (densité) de propagation de la maladie	Taux	Temps ⁻¹	0–∞ →

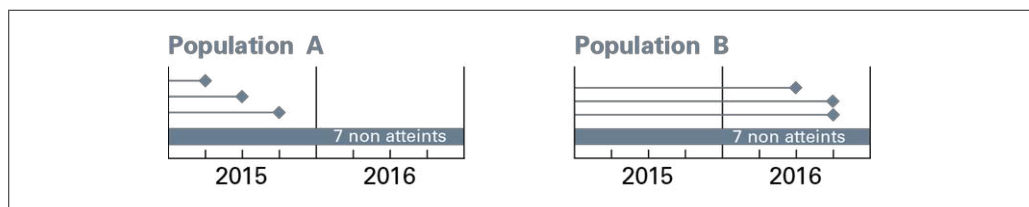
TABLEAU 3.1 ■ Les caractéristiques des mesures utilisées (suite)

Mesure	Usage	Type de rapport	Unité	Domaine de variation
Incidence cumulée	Risque	Proportion	Aucune	0–1
Cote	Estimation du risque	Ratio	Aucune	0–∞

3.2.5 Une synthèse sur les mesures en épidémiologie

Le taux d'incidence et l'incidence cumulée

Supposons que l'on étudie l'occurrence d'un type de cancer dans deux populations constituées chacune de 10 personnes suivies du 1^{er} janvier 2015 au 31 décembre 2016. Supposons aussi que dans chacune d'elles, trois cas de cancer ont été diagnostiqués à des moments différents du suivi. Il reste donc 7 personnes qui ont été suivies jusqu'au 31 décembre 2016 sans avoir été atteintes d'un cancer dans chacune des populations. Les observations donnent lieu au diagramme de la figure 3.2. Dans la population A, le premier cas est apparu à trois mois de suivi, un autre à six mois et le dernier à neuf mois. Dans la population B, le premier cas est apparu à 18 mois de suivi et les deux autres à 21 mois. On calcule, dans les deux cas, une incidence cumulée de 30% (mesure du risque de cancer sur 2 ans). Par ailleurs, visuellement, on constate que la «vitesse» d'apparition des cancers dans la population A est beaucoup plus rapide, ce qui donne lieu à des taux d'incidence différents. Dans la population A, le taux d'incidence est de 3 cas/15,5 personnes-année comparativement à 3 cas/19 personnes-année pour la population B. Le nombre de cas étant le même et le nombre de personnes-année étant plus élevé dans le groupe B, on comprend que la vitesse de survenue est plus importante dans le groupe A.

FIGURE 3.2 ■ Un exemple de diagramme d'observations

3.2.6 La population statique ou dynamique, ouverte ou fermée

Une **population statique** est celle qu'on fige dans le temps pour l'observer. L'étude sur la prévalence du VIH dans l'établissement collégial que fréquente Mirabelle s'est déroulée dans une population statique. Les données ont été recueillies à un moment bien déterminé, comme si l'on faisait une photographie, un instantané de la situation. Une population statique peut donc fournir des proportions, des ratios ou des indices, mais elle ne permet pas de calculer des taux ni des mesures d'incidence. En somme, dans une population statique, on ne peut mesurer ni taux d'incidence ni

incidence cumulée, puisque le temps s'y trouve en quelque sorte suspendu et que l'étude ne tient nullement compte de la durée.

Par contre, une population observée pendant une période plus ou moins longue est dite **population dynamique**. L'étude de Jean sur les infections opportunistes chez les personnes atteintes du sida et celle de Simone sur le VPH sont des exemples d'études portant sur des populations dynamiques. Celles-ci changent avec le temps. Les personnes tombent malades ou guérissent. C'est dans une population dynamique que l'on peut mesurer des taux d'incidence et des incidences cumulées. On peut aussi y mesurer des prévalences si l'on prend des mesures à des moments précis de la période d'observation.

Une population dynamique peut être fermée ou ouverte. L'étude de Simone portant sur le VPH a été menée sur une **population dynamique fermée**. Les étudiantes ont commencé à participer à l'étude au même moment : aucune « nouvelle femme » n'est venue s'y ajouter en cours de route. Elles ont été suivies du début à la fin de la période d'observation. En outre, aucune étudiante n'a « disparu » en cours d'étude. Une population est dite « dynamique fermée » lorsqu'elle présente ces caractéristiques : tous ses membres sont bien identifiés au début de l'étude, personne ne s'ajoute au cours de celle-ci, et chacun est suivi jusqu'à la survenue de l'événement étudié (maladie, décès, accouchement, etc.) ou jusqu'à la fin de la période d'observation. Pour calculer l'incidence cumulée aussi simplement que Simone a pu le faire, il faut se situer dans le contexte d'une population dynamique fermée.

Si l'on s'intéressait à la mortalité dans une ville comme Québec, on considérerait que la population est certes dynamique : cependant, on aurait à tenir compte non seulement des décès, mais aussi des naissances, des emménagements et des déménagements. Il s'agit alors d'une **population dynamique ouverte**. L'incidence cumulée ne peut pas être calculée simplement dans une population ouverte, parce que le nombre de personnes à risque est modifié tout au long de l'étude en fonction des personnes qui s'y ajoutent ou s'en retirent. L'étude de Jean portant sur les infections opportunistes chez les personnes atteintes du sida constitue également un exemple d'étude effectuée sur une population dynamique ouverte.

Dans une population dynamique ouverte stable, lorsque certaines conditions sont réunies, notamment lorsque la prévalence est faible ($< 10\%$), il existe une relation linéaire entre la prévalence (P), le taux d'incidence (I) et la durée moyenne (D) d'une maladie, laquelle s'exprime comme suit :

$$P = I \times D$$

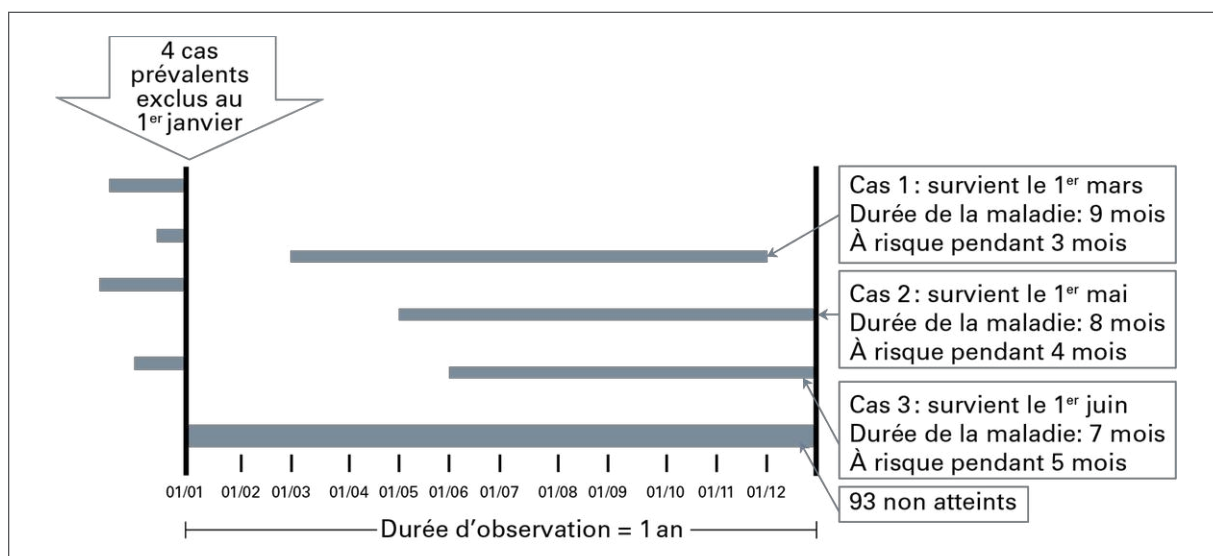
EXEMPLE 3.9

La population des hommes sexuellement actifs est stable dans la région de Montréal. Supposons que la durée de la gonococcie soit de 5 jours et que le taux d'incidence soit de $0,002 \text{ jour}^{-1}$ en 1995. Si la prévalence de la gonococcie était mesurée à un moment quelconque en 1995, elle devrait être de $0,002 \text{ jour}^{-1} \times 5 \text{ jours}$, soit 0,01 ou 1 %.

Une synthèse sur le calcul des mesures dans une population dynamique fermée

Nous l'avons vu, une population dynamique fermée permet de réaliser un calcul plus précis des mesures de fréquence de la maladie qu'une population dynamique ouverte parce qu'une population dynamique fermée favorise un suivi plus précis des personnes observées. La figure 3.3 présente une synthèse sur le calcul des différentes mesures de fréquence dans une population dynamique fermée composée de 100 personnes sélectionnées le 1^{er} janvier 2015 pour être suivies jusqu'au 31 décembre 2015.

FIGURE 3.3 ■ Une synthèse sur le calcul des mesures de fréquence dans une population dynamique fermée



- Calcul de la prévalence au 1^{er} janvier : 4 cas prévalents/100 sujets sélectionnés
- Calcul de l'incidence cumulée : 3 nouveaux cas/96 sujets non atteints au début de la période
- Calcul du taux d'incidence : 3 nouveaux cas/94 personnes-année cumulées à risque
 - Années cumulées à risque par les nouveaux cas : 3 + 4 + 5 mois = 12 mois ou 1 an
 - Années cumulées à risque par les sujets non atteints : 93 × 1 an = 93 années

3.3 Des mesures de mortalité et de morbidité

3.3.1 La mortalité

Le **taux de mortalité** dans une population dynamique fermée se calcule de la même façon que le taux d'incidence, à la différence que des décès figurent au numérateur plutôt que de nouveaux cas de la maladie étudiée. On peut le calculer pour toutes les causes de décès confondues ou pour une ou plusieurs causes précises. Dans une population dynamique ouverte, certaines particularités doivent être prises en compte.

Prenons le taux de mortalité de $7,5 \times 10^{-3} \text{ an}^{-1}$ au Canada en 2015⁴. Puisqu'il s'agit d'un taux, c'est la vitesse avec laquelle la mort s'est propagée dans la population canadienne, tous âges confondus et toutes causes confondues, pendant l'année 2015. La population canadienne constitue une population dynamique ouverte. On ne peut donc pas effectuer un suivi de chaque personne décédée pour délimiter la période de temps écoulée entre le début de la période d'observation (alors que la personne est encore en vie) et le moment de sa mort. Puisqu'on ne connaît pas le moment précis où chaque personne est décédée, on estime les personnes-temps du dénominateur en postulant que les personnes décédées au cours de l'année d'observation ont vécu en moyenne la moitié de la période d'observation qui est ici d'une année. On ne peut pas, non plus, contrôler l'ajout de nouveaux sujets durant la période d'observation. On supposera alors que les nouveau-nés et les immigrants qui sont venus grossir les rangs du groupe en cours de route auront fait partie de celui-ci en moyenne une demi-année. Le taux annuel de mortalité dans de grandes populations est donc généralement obtenu en divisant le nombre de décès enregistrés au cours de l'année par la population au 1^{er} juillet (qui sert d'estimateur du nombre de personnes-temps).

EXEMPLE 3.10

La population de Panacetown était de 608 254 personnes au 1^{er} janvier 2016. Pendant l'année, on y a enregistré 10 402 naissances et 5 201 personnes s'y sont nouvellement installées. Pendant cette même période, 3 211 personnes ont quitté la ville, alors que 6 130 sont décédées. Le taux de mortalité en 2016 a donc été de:

$$\frac{6130}{608\,254 + 1/2(10\,402) + 1/2(5\,201) - 1/2(3\,211) - 1/2(6\,130)}$$

soit à peu près $0,01 \text{ an}^{-1}$ ou $1 \times 10^{-2} \text{ an}^{-1}$ ou 1 % par année.

La probabilité de décès

On peut aussi calculer la probabilité de décès (incidence cumulée) dans une cohorte fermée en rapportant le nombre de décès survenus pendant la période d'étude sur le nombre de personnes à risque en début de suivi. Cet indicateur est utile pour mesurer le risque de décès et sert, entre autres, au calcul de l'espérance de vie.

On trouve dans les recueils statistiques de plusieurs organismes, comme l'Organisation mondiale de la Santé (OMS), le Programme des Nations Unies pour le développement (PNUD), la Banque mondiale, les ministères de la Santé, etc., une série de mesures de l'état de santé appelées « indicateurs ». Parmi les indicateurs les plus utilisés figurent notamment le taux de natalité, le taux de **mortalité néonatale**, le taux de mortalité infantile et le taux de mortalité maternelle. Certains de ces indicateurs, bien qu'ils portent le nom de « taux », sont en fait des ratios, des proportions ou des indices. Les appellations ne sont pas normalisées. Il faut donc se montrer attentif à la signification réelle des mesures présentées.

La létalité

Un autre exemple d'utilisation erronée du mot «taux» est l'expression «taux de létalité». La **létalité** (L), comme on devrait simplement l'appeler, est la proportion des personnes atteintes d'une maladie (m) qui décèdent (d) après un certain temps d'observation. C'est donc une incidence cumulée, qui s'exprime de la façon suivante :

$$L = d/m$$

Un exemple de calcul de la létalité

Au cours d'une étude sur l'infection par le VIH chez les UDI dans le Bronx, on a suivi 62 personnes séropositives pour évaluer la létalité de l'infection à 2 ans, 5 ans et 10 ans. La population étudiée était donc dynamique fermée. En effet, dès qu'une personne devenait séropositive, elle entrait dans l'étude. Bien qu'il y ait eu des entrées pendant deux ans dans cette population, on peut considérer que tous les sujets sont entrés à un moment précis, le même pour tous, c'est-à-dire au moment de leur identification comme séropositifs. À partir de ce moment, chaque personne a été suivie pendant 10 ans ou jusqu'à son décès. Au départ, on observait 62 personnes. Après 2 ans, 12 étaient décédées. Ces personnes n'étaient pas nécessairement décédées du sida. La létalité (toutes causes confondues) à 2 ans chez les UDI séropositifs du Bronx était donc de 12/62 ou 19%. Après 5 ans, 39 personnes étaient décédées. La létalité à 5 ans était donc de 39/62 ou 63%. Après 10 ans, une seule personne vivait encore. La létalité à 10 ans dans ce groupe était ainsi de 61/62 ou 98%.

Dans la formule servant au calcul de la létalité, les décès qui figurent au numérateur sont la somme de tous ceux qui sont survenus dans le groupe suivi, et non uniquement les décès attribuables à la maladie. Puisqu'il s'agit d'une incidence cumulée, la létalité peut aussi s'interpréter comme étant le risque qu'encourt une personne nouvellement atteinte de la maladie considérée de mourir à l'intérieur d'une période donnée. Dans l'étude qui nous occupe, un UDI vivant dans le Bronx qui devient séropositif a 2 chances sur 3 de mourir avant 5 ans et 2 chances sur 100 d'être encore en vie après 10 ans.

3.3.2 La morbidité

Une situation illustrant une mauvaise utilisation du mot «taux» se présente quand les épidémiologistes, confrontés à un épisode aigu comme une intoxication alimentaire, sont appelés à mesurer un **taux d'attaque (TA)**. Celui-ci est défini comme étant la proportion des personnes malades par rapport aux personnes exposées à un risque reconnu. En fait, il s'agit d'une incidence cumulée.

Un exemple du calcul du taux d'attaque

Quelques heures après la fête de Noël célébrée dans une petite usine de quartier, trois personnes se présentent à l'hôpital où Lisa est de garde. Elles ressentent de violentes crampes abdominales, ont des vomissements ainsi que des céphalées et paraissent

épuisées. Au cours de la nuit, Lisa voit arriver 15 autres convives. Au petit matin, la plupart de ces gens sont suffisamment rétablis pour pouvoir rentrer à la maison. Lisa ne veut pourtant pas aller se coucher avant d'avoir mené sa petite enquête pour déceler la cause de cette épidémie. Le tableau 3.2 résume les observations de Lisa parmi les 24 participants au repas.

TABLEAU 3.2 ■ Un exemple de calcul du taux d'attaque

Aliment	A consommé			N'a pas consommé		
	Malade	Non malade	Taux d'attaque*	Malade	Non malade	Taux d'attaque*
Soupe	12	4	12/16 = 75 %	6	2	6/8 = 75 %
Dinde	17	5	17/22 = 77 %	1	1	1/2 = 50 %
Tarte à la crème	18	1	18/19 = 95 %	-	5	0 %

*Incidences cumulées

Quand les inspecteurs du ministère de l'Agriculture, des Pêcheries et de l'Alimentation (inspection des aliments) sont arrivés, Lisa était toute fière de leur annoncer que la tarte était fort probablement l'aliment à l'origine de l'intoxication.

3.3.3 L'espérance de vie

EXEMPLE 3.11

Au Mali, un grand pays du Sahel, l'**espérance de vie** à la naissance était de 57,9 ans en 2014⁵. Au Canada, la même année, l'espérance de vie à la naissance était de 81,9 ans, soit de 80,2 ans pour les hommes et de 84,1 ans pour les femmes.

L'espérance de vie à la naissance se révèle un précieux indicateur de la santé d'une population. On peut la comprendre comme étant l'âge moyen au décès des personnes dans cette population. On pense généralement que plus l'espérance de vie est élevée, plus la santé de la population est bonne. Cet indicateur étant une moyenne, les valeurs extrêmes influent considérablement sur lui. Si, par exemple, beaucoup d'enfants meurent dans les cinq premières années de leur vie, l'espérance de vie à la naissance sera plus faible. Inversement, dans une population où les personnes âgées sont nombreuses, l'espérance de vie à la naissance s'avérera plus élevée.

L'espérance de vie est calculée à partir de la table de mortalité (table de survie). Le tableau 3.3 présente une table abrégée par tranche de 5 années d'âge.

L'espérance de vie pour chaque âge correspond au cumul des années vécues dans l'intervalle d'âge considéré et dans tous les intervalles d'âge subséquents divisé par le nombre de vivants à cet âge. Ce calcul pour la strate d'âge 0 donne l'espérance de vie

TABLEAU 3.3 ■ Une table de mortalité chez les femmes au Québec, 2013-2015⁶

Âge	Survivants	Décès	Probabilité de décès	Années vécues	Espérance de vie
x	l_x	d_x	q_x	L_x	e_x
0	100 000	442	0,00442	99 602	84,1
1	99 558	66	0,00066	398 073	83,5
5	99 493	45	0,00045	497 364	79,5
10	99 448	55	0,00055	497 123	74,6
15	99 393	106	0,00107	496 720	69,6
20	99 287	108	0,00109	496 168	64,7
25	99 179	131	0,00132	495 575	59,7
30	99 048	184	0,00186	494 795	54,8
35	98 864	224	0,00227	493 800	49,9
40	98 640	402	0,00407	492 299	45,0
45	98 238	683	0,00695	489 631	40,2
50	97 555	1 175	0,01204	485 102	35,5
55	96 380	1 872	0,01943	477 568	30,9
60	94 508	2 865	0,03032	465 754	26,4
65	91 643	4 005	0,04370	448 685	22,2
70	87 638	6 184	0,07056	423 570	18,1
75	81 454	9 249	0,11355	385 729	14,2
80	72 206	14 638	0,20272	327 018	10,7
85	57 568	19 370	0,33647	241 349	7,8
90	38 198	20 302	0,53151	138 816	5,4
95	17 896	12 934	0,72272	53 233	3,7
100	4 962	4 962	1,00000	12 767	2,6

x : âge

l_x : survivants à l'anniversaire x

d_x : décès entre les anniversaires x et x + a, a = 1, 4 et 5

q_x : probabilité de décès entre les anniversaires x et x + a, a=1, 4 et 5

L_x : années vécues entre les anniversaires x et x+a, a = 1, 4 et 5

e_x : espérance de vie à l'âge x, c'est-à-dire nombre moyen d'années qu'il reste à vivre à l'anniversaire x

à la naissance, alors que le calcul pour les strates suivantes donne l'espérance de vie à l'âge correspondant à la strate. Par exemple, pour les personnes ayant atteint l'âge de 10 ans, l'espérance de vie présentée dans la colonne de droite (74,6 ans) correspond au nombre moyen d'années qui reste à vivre pour les individus après 10 ans.

L'espérance de vie peut donc être estimée à des âges variés. Par exemple, toujours selon le tableau 3.3, l'espérance de vie à 50 ans, c'est-à-dire le nombre moyen d'années qui reste à vivre pour les femmes ayant atteint l'âge de 50 ans, est de 35,5 ans. De façon générale, l'espérance de vie des femmes et des hommes de 65 ans était respectivement de 22,2 ans et de 19,2 ans en 2014 au Québec (Institut de la statistique du Québec, 2015). Cela signifie que les femmes et les hommes ayant atteint 65 ans ont encore en moyenne respectivement 22,2 et 19,1 années à vivre, et que les femmes de 65 ans peuvent s'attendre à vivre en moyenne 3 ans de plus que les hommes de 65 ans.

On peut aussi estimer l'espérance de vie en bonne santé ou sans incapacité. L'espérance de vie sans limitation d'activité est un indicateur plus large que celui de l'espérance de vie, qui introduit le concept de la qualité de vie. Il permet de distinguer les années de vie libres de toute limitation d'activité des années vécues avec au moins une limitation d'activité. La définition de l'état de santé utilisée pour le calcul des indicateurs d'espérance de vie en santé pourra varier selon les auteurs. Il faudra être attentif à la méthode qu'ils utilisent. On trouvera, selon les auteurs, des mesures de dépendance, d'incapacité, de désavantage, de handicap et de limitations d'activités.

3.4 L'ajustement des mesures

En parcourant la littérature, Lisa a remarqué une chose pour le moins étrange. Panacetown, qu'elle connaît bien pour y être allée il y a quelques années, affiche un taux de mortalité de $10 \times 10^{-3} \text{ an}^{-1}$, alors que Ciudad de Niños présente un taux de mortalité de $9 \times 10^{-3} \text{ an}^{-1}$. Or, Panacetown est une ville cossue, alors que Ciudad de Niños est une jeune ville qui s'est développée de façon un peu anarchique depuis une vingtaine d'années. Lisa se serait donc attendue à ce que la mortalité soit moins élevée à Panacetown.

En fouillant un peu, elle a pu mettre la main sur des données fort intéressantes, réunies dans le tableau 3.4, à la page suivante, où l'on présente les populations des deux villes (en personnes-année) stratifiées selon l'âge, la mortalité pour chaque groupe d'âge et la mortalité globale. La mortalité spécifique, c'est-à-dire la mortalité dans chaque tranche d'âge, confirme pourtant son intuition. Dans chaque groupe d'âge, la mortalité est plus élevée à Ciudad de Niños.

D'où vient ce paradoxe? Si l'on examine la répartition des personnes dans les groupes d'âge, on constate que la structure d'âge des deux populations est bien différente. Les habitants de Ciudad de Niños sont bien plus jeunes que ceux de Panacetown. Cette différence dans la structure d'âge introduit une distorsion dans la mesure globale. Pour comparer ces deux villes correctement, il faut donc observer les mesures spécifiques, c'est-à-dire comparer les deux populations, tranche d'âge par tranche d'âge.

Pour exprimer cette différence en un seul rapport, il suffit d'ajuster les mesures globales en faisant le raisonnement suivant : si l'on rapporte les taux de mortalité

spécifiques des deux villes à une même population fictive, 1 000 000 d'habitants par exemple, on obtient une mesure globale correcte, puisque l'on compare des données semblables. C'est ce qu'on appelle l'**ajustement** des mesures. En faisant cet exercice, on obtient un portrait semblable à celui du tableau 3.5.

TABLEAU 3.4 ■ Des populations stratifiées selon l'âge, la mortalité spécifique et la mortalité globale

Ville	Groupe d'âge (ans)	Effectif (personnes-année)	%	Décès	Mortalité/1 000
Panacetown	0-20	117 246	19	1 053	8,98
	21-40	120 579	20	824	6,83
	41-60	185 124	30	1 024	5,53
	> 60	188 536	31	3 229	17,13
	Total	611 485	100	6 130	10,02
Ciudad de Niños	0-20	317 246	48	3 073	9,69
	21-40	195 579	30	1 544	7,89
	41-60	120 124	18	806	6,71
	> 60	25 501	4	507	19,88
	Total	658 450	100	5 930	9,01

TABLEAU 3.5 ■ L'ajustement des taux sur une population fictive

Ville	Groupe d'âge (ans)	Effectif	%	Décès	Mortalité/1 000
Panacetown	0-20	250 000	25	2 245	8,98
	21-40	250 000	25	1 708	6,83
	41-60	250 000	25	1 383	5,53
	> 60	250 000	25	4 282	17,13
	Total	1 000 000	100	9 618	9,62
Ciudad de Niños	0-20	250 000	25	2 422	9,69
	21-40	250 000	25	1 974	7,89
	41-60	250 000	25	1 677	6,71
	> 60	250 000	25	4 970	19,88
	Total	1 000 000	100	11 043	11,04

Après ajustement, la mortalité est effectivement plus élevée à Ciudad de Niños ($11 \times 10^{-3} \text{ an}^{-1}$) qu'à Panacetown ($10 \times 10^{-3} \text{ an}^{-1}$). Comment a-t-on fait cet ajustement? On a d'abord appliqué les taux de décès spécifiques de chaque ville à l'effectif de la population fictive pour trouver le nombre de décès attendus. Par exemple, à Panacetown, on a un taux de décès de 0,008 98 pour la tranche d'âge de 0-20 ans : en effet, $250\,000 \times 0,008\,98 = 2\,245$. À Ciudad de Niños, on a un taux de décès de 0,019 88 pour la tranche d'âge de 61 ans : en effet, $250\,000 \times 0,019\,88 = 4\,970$. Ensuite, en additionnant les décès ainsi obtenus et en faisant le rapport à la population fictive totale, on obtient un taux ajusté. Cette méthode s'appelle l'**ajustement direct** (ou standardisation directe). On s'est servi des taux spécifiques réels qu'on a appliqués à une distribution de population identique fictive. Ces taux fictifs ajustés par standardisation permettent maintenant de comparer la mortalité dans les deux populations sans l'effet confondant de l'âge. L'ajustement des mesures peut également être effectué par l'**ajustement indirect**, une méthode qui consiste, à l'inverse, à appliquer des taux identiques fictifs aux distributions de population réelles.

3.5 Les modèles (ou types) d'études en épidémiologie

Au ministère, Jean est chargé de l'étude et du contrôle des infections transmissibles sexuellement et par le sang (ITSS). Il doit pouvoir répondre à des questions comme celles-ci : « Quelles sont les caractéristiques des personnes qui sont atteintes de gonococcie dans la province? », « La distribution gratuite de condoms dans les écoles, les entreprises, les bars, les hôtels et autres lieux de rencontre influe-t-elle sur l'incidence des ITSS? », « Les personnes atteintes d'une ITSS ont-elles eu un nombre de partenaires sexuels plus élevé que les autres au cours de la dernière année? »

Toutes ces questions, de même que toutes celles posées dans le domaine de la santé, peuvent se résumer à deux interrogations fondamentales, soit « quoi? » et « pourquoi? ». Dans le premier cas, on s'applique à décrire un phénomène, et dans le second, on cherche les causes d'un phénomène.

L'épidémiologie offre à Jean trois méthodes de base pour mener ses études : l'étude transversale, l'étude longitudinale et l'étude cas-témoins. Le choix de l'une ou l'autre dépendra des questions auxquelles son étude devra répondre et des moyens financiers dont il disposera.

3.5.1 L'étude transversale

Dans l'**étude transversale**, les données sont recueillies au cours d'une enquête menée à un moment précis dans le temps. Il ne se fait pas de suivi. L'étude transversale permet de décrire un problème de santé à un temps donné. Elle produit une photographie de la situation au moment de l'étude, dans la population étudiée. La prévalence est la mesure obtenue à la suite de l'étude transversale. Une étude transversale pourrait répondre à la question : « Quelles sont les caractéristiques des personnes qui sont atteintes de gonococcie dans la province? » Elle ne permettrait cependant pas de répondre aux autres questions auxquelles Jean se trouve confronté.

3.5.2 L'étude longitudinale

Alors que l'étude transversale « fige » le temps, l'**étude longitudinale** se déroule dans le temps et permet l'observation de l'évolution des phénomènes. Cette méthode s'applique à l'étude d'un cas (observation d'une personne), d'une série de cas (observation de plusieurs personnes dispersées dans le temps et l'espace) ou d'une ou de plusieurs populations. Quand l'étude longitudinale porte sur une ou plusieurs populations, on l'appelle « **étude de cohorte(s)** ».

Une cohorte est un ensemble d'individus ayant vécu une situation semblable au cours d'une même période. Ainsi, il y a 2 000 ans, les soldats romains, recrutés à un moment donné, portaient tous ensemble pour conquérir de nouveaux territoires. Pendant toute leur affectation, ils n'accueillaient aucune recrue dans leurs rangs. Le groupe ainsi constitué s'appelait une « cohorte ». À la fin de l'affectation, la cohorte revenait à Rome, amputée des soldats qui étaient décédés ou qui avaient déserté. Par analogie, les études de cohorte(s) se définissent par l'étude d'un ou de quelques facteurs susceptibles de provoquer l'apparition d'une maladie dans une ou des populations (classiquement, elles seront fermées, mais par extension, à certaines conditions, elles pourront être ouvertes).

À partir de ce type d'étude, on peut donc décrire un phénomène qui se déroule dans le temps. On peut aussi estimer la fréquence de la maladie chez des sujets exposés à un facteur quelconque et la fréquence de cette maladie chez des sujets qui n'y sont pas exposés. Puisque l'observation porte sur chacun des membres d'une cohorte, il est possible de compter le nombre de nouveaux cas de la maladie. On peut également mesurer le temps écoulé entre le début de l'observation et l'apparition de la maladie, le décès ou la fin de la période d'observation. Les études de cohorte(s) permettent donc de recueillir des mesures comme le taux d'incidence et l'incidence cumulée, ainsi que d'obtenir des prévalences et des cotes. C'est le modèle d'étude idéal en épidémiologie. Une étude de cohorte(s) permettrait de répondre à toutes les questions auxquelles Jean s'intéresse.

La personne qui mène l'étude de cohorte(s) connaît le ou les facteurs d'exposition des personnes qui seront suivies. Elle suit dans le temps les personnes exposées ou non pour voir si elles contracteront la maladie. On dit de cette étude qu'elle est « prospective », puisque l'exposition mesurée ou constatée coïncide avec le début de l'étude. Il est aussi possible de faire des études de cohorte(s) rétrospectives (ou historiques) si l'étude se déroule à partir de personnes dont on connaît l'exposition à des événements documentés dans le passé et qu'il est possible d'établir la relation entre l'exposition et l'état de santé plus tard dans le temps. Par exemple, une étude longitudinale permettrait de répondre à la question : « La distribution gratuite de condoms dans les écoles, les entreprises, les bars, les hôtels et autres lieux de rencontre influe-t-elle sur l'incidence des ITSS ? »

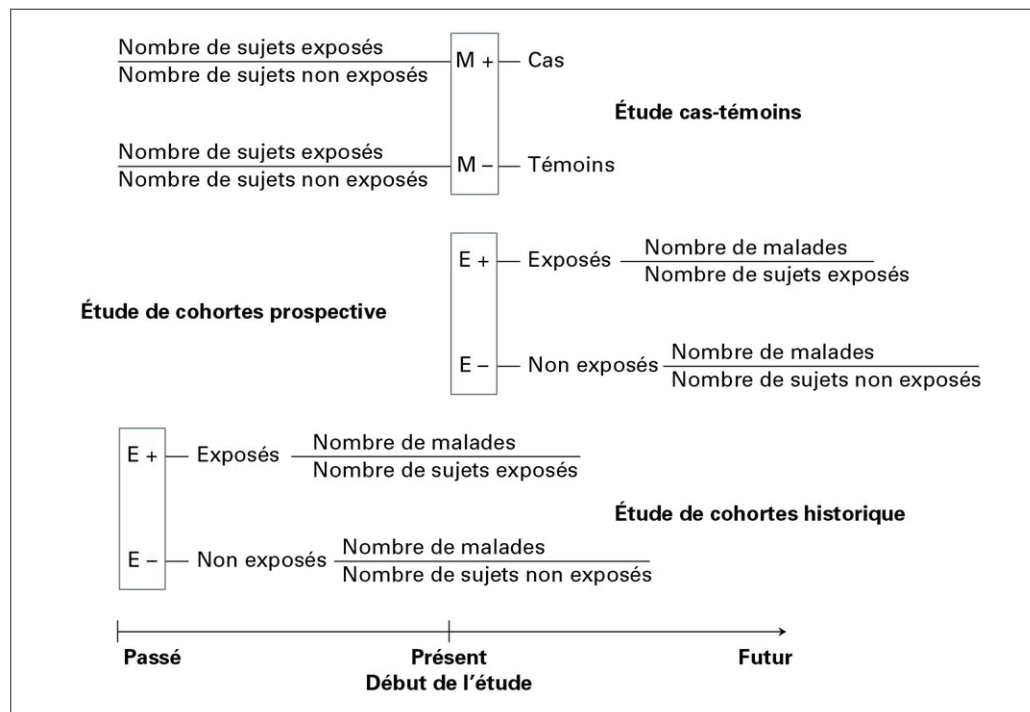
3.5.3 L'étude cas-témoins

Comme son nom l'indique, l'**étude cas-témoins** implique d'emblée la comparaison de deux groupes : l'un composé de personnes atteintes de la maladie étudiée, appelées les « cas », et l'autre formé de personnes n'ayant pas contracté la maladie, appelées les « témoins ». La maladie sert en quelque sorte de point de départ ; on cherche ensuite une exposition antérieure. En d'autres termes, on connaît des personnes malades, les cas, et l'on connaît des personnes qui auraient pu être malades, mais qui ne le sont pas, les témoins. À partir de ces données, on cherche le ou les facteurs qui ont conduit les uns à être malades et qui ont épargné les autres. Dans une étude cas-témoins, l'exposition mesurée ou constatée est antérieure au début de l'étude.

Les études cas-témoins sont généralement utilisées lorsque la maladie est rare ou lorsque le facteur d'exposition demeure inconnu. La cote de l'exposition chez les cas et celle de l'exposition chez les témoins sont les mesures obtenues à la suite de ce type d'études. Leur ratio constitue le rapport de cotes (communément appelé le *odds ratio*). Par exemple, une étude cas-témoins pourrait répondre à la question : « Les personnes atteintes d'une ITSS ont-elles eu un nombre de partenaires sexuels plus élevé que les autres au cours de la dernière année ? »

La figure 3.4 illustre ce concept et indique aussi les rapports qui sont mesurés dans les divers groupes comparés.

FIGURE 3.4 ■ La relation temporelle entre certains types d'études épidémiologiques



3.6 Une classification des études épidémiologiques

La classification des études épidémiologiques pourra varier de façon importante selon les auteurs. Il n'existe pas de consensus sur cette question. La classification présentée ici s'appuie sur les grands principes et sur les distinctions fondamentales que l'on trouvera dans la littérature en sciences de la santé.

3.6.1 Les études à visée étiologique

Un exemple d'étude transversale

Une nouvelle usine de production de plastique est installée depuis plus d'un an dans le voisinage de la polyclinique Milo. La population du quartier s'est plainte des odeurs et de malaises divers. L'équipe de la polyclinique émet l'hypothèse que les émanations de l'usine peuvent être à l'origine des malaises. Pour appuyer ou infirmer cette hypothèse, elle décide de mener une enquête sur l'état de santé des voisins de l'usine pour le comparer à celui des habitants d'une autre partie du quartier. Comme l'équipe ne dispose pas de gros budgets, l'enquête sera menée en deux semaines et portera surtout sur les problèmes de santé actuels. Il s'agit donc d'une étude à visée étiologique, puisqu'on utilise un groupe de comparaison; elle est transversale, puisqu'il n'y a pas de suivi et qu'elle porte sur des cas prévalents.

Les études longitudinales

L'étude de cohorte(s) expérimentale L'étude de type expérimental se rapproche des expériences scientifiques couramment pratiquées en chimie, en physique ou dans les autres sciences de la nature. Toutes les variables de temps, de lieux et de personnes sont contrôlées. Dans ces conditions, les chercheurs peuvent observer les variations de la maladie en fonction des variations de l'exposition. Ce type d'étude doit être reproductible par tout autre chercheur.

Dans le domaine de la santé humaine, il est difficile de mener des études expérimentales, surtout pour des raisons éthiques. Par exemple, il serait inacceptable d'inoculer une bactérie à des personnes choisies au hasard pour observer l'apparition éventuelle d'une maladie. Par contre, une compagnie pharmaceutique pourrait demander à Jean de l'aider à évaluer un nouveau médicament contre la gonococcie. Jean pourrait alors choisir au hasard un groupe de personnes parmi celles déjà atteintes de la maladie. Toujours au hasard, il administrerait le nouveau médicament à une moitié du groupe et le médicament habituel à l'autre moitié. En comparant les résultats obtenus dans les deux groupes, il pourrait évaluer l'efficacité du nouveau médicament. Cette étude de cohortes expérimentale constitue ce que l'on appelle un « essai thérapeutique hasardisé » (certains utilisent l'anglicisme « randomisé »).

Les études expérimentales sont particulièrement utiles pour déterminer l'efficacité d'un agent ou d'une intervention thérapeutique, établir l'efficacité d'une intervention ou d'un programme préventif et évaluer l'efficacité de différentes méthodes de distribution des services ou des soins de santé au sein d'une population. Dans ce genre d'études, pour éviter les biais (ou erreurs systématiques, une notion abordée plus loin dans le présent ouvrage), on procède souvent en faisant une **affectation à l'aveugle** (le chercheur et le patient ne sachant pas ni l'un ni l'autre à quel groupe est affecté le patient) et une **appréciation à l'aveugle** (la personne qui apprécie ou mesure le résultat ignorant quel traitement a reçu le patient).

Lorsque l'une ou l'autre de ces techniques est utilisée, on dit que l'on mène une étude en **simple aveugle**. L'étude est dite en **double aveugle**, ou à **double insu**, lorsqu'elle comprend à la fois une affectation et une appréciation à l'aveugle. C'est l'idéal parce que l'on est alors certain que les attentes des chercheurs ou des patients ne viennent pas fausser les résultats. Toutefois, ce n'est pas toujours possible. Quand le traitement implique une intervention chirurgicale ou le port d'un appareil, par exemple, l'étude ne peut habituellement pas être menée en double aveugle.

Un exemple d'étude de cohorte(s) non expérimentale Comme il n'est pas toujours possible de mener des études expérimentales, on se limite souvent à l'observation des phénomènes tels qu'ils se produisent naturellement. Si Jean voulait connaître le degré d'efficacité du condom contre les ITSS, il ne pourrait certes pas choisir des gens et exiger que, d'une part, la moitié d'entre eux, sélectionnés au hasard, portent le condom et que, d'autre part, l'autre moitié de ces gens ne le portent pas. Il devrait plutôt mener une enquête afin de recruter des personnes qui utilisent déjà ce préservatif et comparer l'incidence des ITSS dans cette cohorte à celle des ITSS dans la cohorte de gens qui ne l'utilisent pas. C'est ce que l'on appelle une étude de type «non expérimental» ou «d'observation».

Un exemple d'étude de cohorte(s) quasi expérimentale Jean se demande aussi si la distribution gratuite de condoms dans les écoles, les entreprises, les bars, les hôtels et autres lieux de rencontre influe sur l'incidence des ITSS. La meilleure façon de répondre à cette question est d'aller vérifier sur le terrain. Pour que ce soit faisable, Jean commence par limiter son champ d'action aux écoles. Il repère justement deux collèges qui accueillent à peu près le même type de clientèle dans deux villes comparables du point de vue sociodémographique. Jean propose un programme de prévention des ITSS dans les deux établissements. Dans l'un et l'autre cas, on commencera par une surveillance de quelques mois qui permettra d'établir une comparaison des incidences des ITSS. Puis, on entamera le même programme de prévention dans les deux collèges. Par contre, dans l'un d'entre eux, on distribuera gratuitement les condoms, alors que, dans l'autre, ils seront vendus au prix courant.

Ici, Jean contrôle les variables de lieux et de temps : il décide quand et où seront distribués les condoms. Cependant, il ne contrôle pas vraiment les variables de personnes puisqu'il ne peut choisir au hasard celles qui paieront leurs condoms. Cela est déjà déterminé par le collège qu'elles fréquentent. Une étude quasi expérimentale est une étude dans laquelle le chercheur contrôle l'ensemble des facteurs, sauf les variables de personnes.

Un exemple d'étude cas-témoins Jean se demande si les personnes atteintes d'une ITSS ont eu un nombre plus élevé de partenaires sexuels que les autres au cours de la dernière année. Il choisit de mener une étude cas-témoins. Après un calcul savant, il estime qu'il lui faut comparer 25 personnes qui ont consulté pour une ITSS à la polyclinique Milo (cas) à 100 personnes qu'il recrutera dans la même clinique, mais qui auront consulté pour autre chose qu'une ITSS (témoins). Toutes ces personnes répondront à un même questionnaire. D'emblée, l'étude cas-témoins est une étude à visée étiologique, puisque l'on compare toujours deux groupes, les malades et les témoins, au regard d'un certain nombre de variables qui seraient susceptibles d'expliquer pourquoi les uns sont atteints de la maladie alors que les autres ne le sont pas.

Les avantages et les désavantages des études de cohorte(s) et cas-témoins

Dans le meilleur des mondes, on devrait toujours mener des études de cohorte(s) expérimentales. Toutefois, on a vu que, pour des considérations éthiques et pratiques, les études expérimentales sont souvent impossibles. Par ailleurs, l'ensemble des moyens nécessairement mis en œuvre pour réaliser une étude de cohorte(s) qui s'étendra sur de longues années coûte très cher. Pour des raisons économiques, on optera donc souvent pour les études cas-témoins.

Les études de cohorte(s) et les études cas-témoins comportent en fait autant d'avantages que de désavantages, respectivement exposés dans le tableau 3.6.

TABLEAU 3.6 ■ Une comparaison des avantages et des désavantages des études de cohorte(s) et des études cas-témoins

Études de cohorte(s)	Études cas-témoins
Elles sont généralement plus longues et plus coûteuses.	Elles sont généralement moins longues et moins coûteuses.
La connaissance de l'exposition peut influencer le diagnostic.	Le diagnostic est déjà posé.
L'information obtenue sur l'exposition ne peut être biaisée par la connaissance de la maladie.	La connaissance de la maladie peut influencer la recherche de l'exposition.
Le risque d'une perte au suivi de sujets demeure toujours présent.	Ce risque est absent.

TABLEAU 3.6 ■ Une comparaison des avantages et des désavantages des études de cohorte(s) et des études cas-témoins (suite)

Études de cohorte(s)	Études cas-témoins
Le risque de biais est moins important.	Le risque de biais est plus important.
Elles sont inefficaces lorsque la maladie est rare.	Elles sont utiles lorsque la maladie est rare.
Elles sont très utiles lorsque l'exposition est rare.	Elles sont inefficaces lorsque l'exposition est rare.

3.6.2 Les études descriptives

L'étude descriptive de population transversale

Les caractéristiques de l'**étude descriptive transversale** sont qu'elle est effectuée sans groupe de comparaison et à un moment précis dans le temps. Comme les autres études descriptives, elle aboutira à un compte rendu qui permettra d'émettre des hypothèses, mais elle ne fournira pas d'explications.

EXEMPLE 3.12

Jean décide de mener une petite étude pour savoir qui est atteint de gonococcie dans la province. Il demandera à 10 personnes, choisies au hasard parmi celles venant de recevoir un diagnostic de gonococcie dans un CLSC, de remplir un questionnaire anonyme sur leurs habitudes sexuelles et leurs caractéristiques sociales. Cette enquête se tiendra le mois prochain.

L'étude descriptive de population longitudinale

Dans une **étude descriptive longitudinale**, il n'y a pas de groupe de comparaison. On dit qu'elle est longitudinale parce qu'il y a un suivi dans le temps, au sein d'une population dynamique ouverte, au sujet de laquelle les données démographiques (naissances, décès, arrivées et départs) sont suffisamment précises pour que l'on puisse estimer la population à risque et les personnes-temps à risque à tout moment. Enfin, une telle étude est rétrospective ou historique, puisque le moment de survenue du phénomène étudié est bien antérieur à la date du début de l'étude elle-même.

EXEMPLE 3.13

Au ministère de la Santé, des collègues de Jean, alarmés par une recrudescence de la tuberculose dans certaines communautés autochtones, veulent obtenir une description de l'évolution de cette pathologie depuis 10 ans. Comme les données sanitaires et démographiques ont été recueillies consciencieusement dans la communauté inuite depuis de nombreuses années, les épidémiologistes du Ministère ont choisi d'analyser la progression de la tuberculose dans cette communauté au cours de la dernière décennie.

Un exemple d'étude d'un cas ou d'une série de cas

Dans sa pratique à la polyclinique Milo, Lisa en voit de toutes les couleurs. Joël, un travailleur social de la polyclinique, lui demande de rencontrer une patiente qui manifeste un comportement étrange. La patiente, M^{me} Particulier, ne se présente plus au travail et ne répond plus au téléphone. Elle a expliqué au travailleur social que si elle n'était pas allée au travail depuis plusieurs jours, c'est parce qu'elle se sentait grippée, mais aussi parce qu'elle devait rédiger un guide de voyage interplanétaire avant la semaine prochaine. Elle a expliqué que, peu de temps après le début de sa grippe, des extraterrestres lui avaient rendu visite et lui avaient demandé de rédiger ce guide à l'intention de tous ceux qui voudraient venir les rencontrer chez eux.

Lisa a fait subir une série d'examens biochimiques et radiologiques à M^{me} Particulier. Tous les examens se sont révélés à peu près normaux, confirmant tout au plus un syndrome grippal, mais Lisa a décidé de la mettre en observation. Deux jours plus tard, M^{me} Particulier, visiblement anxieuse, a mentionné à Lisa qu'elle n'avait aucun souvenir de ce qui s'était passé depuis les 10 derniers jours et qu'elle croyait être devenue folle. Lisa l'a rassurée du mieux qu'elle a pu, lui a raconté ce qu'elle savait, a refait un examen complet, a consulté des collègues et a suivi sa patiente pendant plus d'un an. Tous les examens se sont révélés normaux, et M^{me} Particulier n'a jamais connu de récurrence. Par la suite, Lisa et Joël ont exploré la littérature sans rien trouver de semblable. Ils ont donc décidé de rédiger un article pour exposer ce cas d'amnésie totale, survenue dans un contexte grippal et dont la guérison a été spontanée, et ils l'ont publié dans le *Journal de l'Association médicale*.

Une telle présentation de cas est fréquente dans la littérature médicale. Si Lisa et Joël avaient traité du même syndrome survenu chez plusieurs personnes, leur article aurait porté sur l'**étude d'une série de cas**. C'est ce type d'études qui a en premier lieu attiré l'attention sur l'émergence de l'épidémie de sida. À la portée de toutes les personnes travaillant dans le domaine de la santé, ces études permettent de rendre compte de phénomènes nouveaux ou jusque-là méconnus. Elles ouvrent la voie à des études descriptives de population et, parce qu'elles permettent la formulation d'hypothèses, elles tracent aussi le chemin à des études à visée étiologique.

L'étude écologique

Il est possible de varier interminablement la combinaison des méthodes, des visées et des plans d'études. En particulier, si l'on recueille des renseignements sur des sous-populations plutôt que sur des personnes, on dira que l'on fait une **étude écologique**. Dans ce type d'étude, l'unité d'observation n'est plus une personne, mais un groupe de personnes: classe, usine, communauté, quartier, ville, etc. Ce type d'étude reste du domaine descriptif. On pourrait par exemple établir dans un graphique la corrélation entre la proportion de femmes qui ont annuellement un dépistage des lésions

du col utérin (test Pap) (axe des x) et le taux d'incidence du cancer du col (axe des y) pour chaque province du Canada. Un tel graphique permettrait de voir la relation entre les deux variables. Théoriquement, les taux les plus faibles de cancer du col utérin devraient être notés dans les provinces où la proportion de femmes qui subissent régulièrement un test Pap est élevée.

Ce type d'étude pourra suggérer des hypothèses qui seront récupérées par des devis de recherche capables d'ouvrir la voie à un traitement plus rigoureux de l'information, comme les études à visée étiologique. On sera donc prudent en évitant de «projeter» une découverte à l'échelle écologique, sur une base individuelle, ce qu'on appelle le «piège écologique» (*ecological fallacy*).

EXEMPLE 3.14

Blanche, une sage-femme, s'intéresse à l'effet des maisons de naissance sur la croissance et le bien-être de la population. Elle choisit de comparer les caractéristiques des communautés dotées d'une maison de naissance à celles des communautés où il n'y en a pas.

L'étude de tendance

Un autre type d'étude souvent utilisé est l'**étude de tendance**, aussi appelée «séries chronologiques», qui vise à mesurer la fréquence et à évaluer la répartition d'un phénomène (maladie ou décès dans le temps). Une façon de mener une étude de tendance consiste à répéter à intervalles réguliers des études transversales auprès des mêmes populations.

EXEMPLE 3.15

Raymond, un gérontologue, veut connaître l'évolution de la grippe dans les établissements de soins de longue durée sur une période de 10 ans. Il choisit cinq centres de la région (dans le jargon du métier, on les appelle les «centres sentinelles»). Tous les ans, en novembre, il mesurera la prévalence de la grippe dans ces centres. L'analyse de ces données périodiques lui permettra de dégager la tendance de la fréquence de cette pathologie dans ce milieu.

3.6.3 Une classification des types d'études selon leur finalité

En somme, comme on l'a vu, les questions que se pose Jean peuvent être classées en deux grandes catégories: d'une part, «quoi?» (la description d'un phénomène) et, d'autre part, «pourquoi?» (la recherche des causes, l'étiologie du phénomène). L'existence d'un groupe de comparaison distingue essentiellement les études à visée étiologique (appelées aussi études analytiques) des études descriptives. Le tableau 3.7 fournit un aperçu succinct d'une classification des divers types d'études en fonction de leur finalité.

TABLEAU 3.7 ■ La classification des types d'études en fonction de leur finalité

Répondre à la question « Quoi ? », décrire un phénomène	
Études descriptives	<ul style="list-style-type: none"> • d'un cas • d'une série de cas • de populations : <ul style="list-style-type: none"> - transversales - longitudinales
Répondre à la question « Pourquoi ? », la recherche des causes, l'étiologie du phénomène	
Études à visée étiologique	<ul style="list-style-type: none"> • transversales • cas-témoins • longitudinales de cohorte(s) : <ul style="list-style-type: none"> - expérimentales - non expérimentales - quasi expérimentales

En résumé

Dans ce chapitre, nous avons passé en revue les outils permettant d'établir les bases du jugement clinique concernant les risques auxquels sont exposés les patients. La démarche clinique s'alimente des données tirées de l'anamnèse de la personne qui consulte ainsi que des données portant sur ses habitudes de vie et sur ses environnements social, professionnel et autres. Quand elles sont chiffrées, les données doivent être relativisées par rapport à d'autres données, ce qui permet d'obtenir une idée plus exacte de leur poids et de saisir l'importance du problème qui fait l'objet d'investigation. L'utilisation du rapport sous différentes formes (proportion, taux, ratio, indice) aide à établir la comparaison, mais une prise en compte des caractéristiques de la population étudiée (statique ou dynamique, ouverte ou fermée) est nécessaire, comme il est aussi essentiel de procéder à des ajustements quand il est question de comparer des données provenant de deux ou de plusieurs populations.

Les ajustements, qui peuvent être obtenus selon une procédure directe ou indirecte, consistent à standardiser les données à partir des mêmes référentiels, permettant ainsi de réduire les effets des facteurs de confusion. La prévalence, le taux d'incidence et l'incidence cumulée sont des mesures de fréquence ou de densité régulièrement utilisées en épidémiologie pour faire état de la manifestation des problèmes de santé et des effets des interventions mises en place pour les contrôler. L'espérance de vie et bien d'autres indicateurs de la santé (mortalité, morbidité, létalité) fournissent des renseignements sur la santé et sur la qualité de vie d'une population. Afin de décrire les problèmes de santé, d'en expliquer l'étiologie et d'évaluer l'efficacité des interventions, l'épidémiologie mène des études en utilisant trois principaux modèles (étude transversale, étude longitudinale, étude cas-témoins) et en s'appuyant sur des devis expérimentaux, quasi expérimentaux ou non expérimentaux.

Exercices**Voir la section complète des exercices sur le Web.**

1. Quelle mesure de fréquence de la maladie doit-on utiliser pour estimer le risque pour un individu d'être atteint de la maladie durant une certaine période d'observation?
 - a) Taux d'incidence.
 - b) Prévalence.
 - c) Incidence cumulée.
 - d) Létalité.
2. Quelle mesure de fréquence de la maladie doit-on utiliser pour obtenir la vitesse d'apparition de la maladie dans une population durant une certaine période d'observation?
 - a) Taux d'incidence.
 - b) Prévalence.
 - c) Incidence cumulée.
 - d) Létalité.
3. On vient de mettre sur le marché un nouveau médicament qui prolonge la vie des personnes atteintes du sida, mais qui ne permet pas de prévenir la survenue de la maladie. Que se passera-t-il?
 - a) Le taux d'incidence du sida diminuera.
 - b) Le taux d'incidence du sida augmentera.
 - c) La prévalence du sida diminuera.
 - d) La prévalence du sida augmentera.
4. Si vous vouliez étudier les facteurs de risque d'une maladie rare comme le cancer du pancréas, quel type d'étude épidémiologique devriez-vous privilégier?
 - a) Une étude à visée étiologique cas-témoins.
 - b) Une étude à visée étiologique de cohorte longitudinale.
5. Vrai ou faux. Une étude transversale permet de calculer un taux d'incidence.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 4

Collaboration à la 3^e édition :

Pierre Deshaies

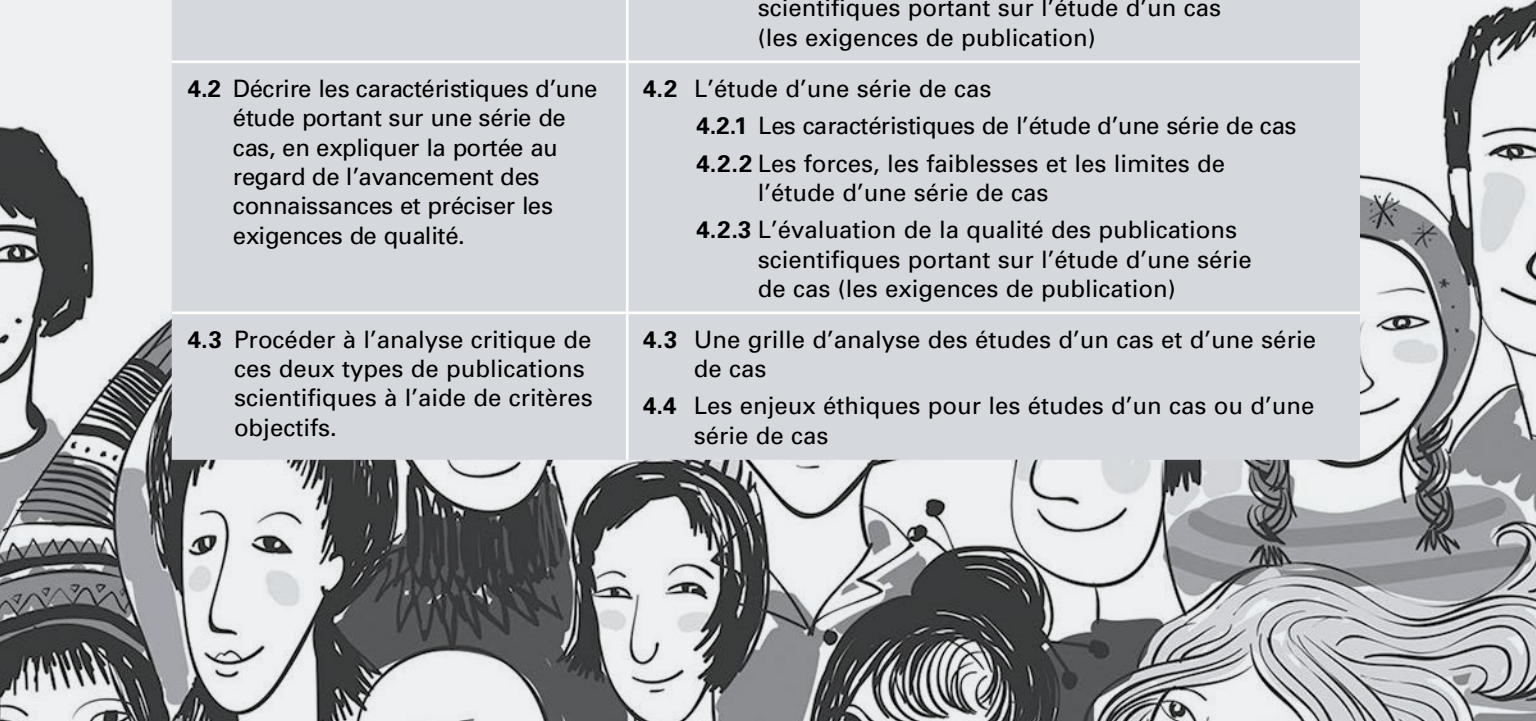
André Simpson

La lecture critique des études descriptives : études d'un cas et d'une série de cas

Au terme de ce chapitre, vous pourrez :

- juger de la pertinence d'un article portant sur un cas ou une série de cas ;
- évaluer la qualité scientifique d'un article portant sur un cas ou une série de cas ;
- soupeser l'utilité, pour votre pratique, du contenu et des résultats d'un article portant sur un cas ou une série de cas ;
- prendre une décision appropriée sur la pertinence de modifier votre pratique clinique à la lumière d'un article portant sur un cas ou une série de cas.

Objectifs spécifiques	Contenu
4.1 Décrire les caractéristiques d'une étude portant sur un cas, en expliquer la portée au regard de l'avancement des connaissances et préciser les exigences de qualité.	4.1 L'étude d'un cas 4.1.1 Les caractéristiques de l'étude d'un cas 4.1.2 Les forces, les faiblesses et les limites de l'étude d'un cas 4.1.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'un cas (les exigences de publication)
4.2 Décrire les caractéristiques d'une étude portant sur une série de cas, en expliquer la portée au regard de l'avancement des connaissances et préciser les exigences de qualité.	4.2 L'étude d'une série de cas 4.2.1 Les caractéristiques de l'étude d'une série de cas 4.2.2 Les forces, les faiblesses et les limites de l'étude d'une série de cas 4.2.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'une série de cas (les exigences de publication)
4.3 Procéder à l'analyse critique de ces deux types de publications scientifiques à l'aide de critères objectifs.	4.3 Une grille d'analyse des études d'un cas et d'une série de cas 4.4 Les enjeux éthiques pour les études d'un cas ou d'une série de cas





Mise en situation

Sarah consulte un jeudi soir à l'urgence de l'hôpital où Lisa, omnipraticienne à la polyclinique Milo, est la médecin de garde. Âgée de 20 ans, étudiante en sciences de la santé, elle lui raconte qu'elle souffre depuis cinq jours de nausées, de vomissements et de douleurs abdominales. Ces malaises sont apparus environ huit heures après l'ingestion de ce qu'elle croit être des champignons « magiques ». Par ailleurs, la jeune femme affirme n'avoir eu aucune hallucination à la suite de leur ingestion.

Sarah n'a aucun antécédent médical ni chirurgical particulier. À l'examen physique, Lisa note une pression artérielle élevée, soit 160/100 mm Hg. La valeur du potassium sérique, soit 5,3 mmol/l, et celle de l'urée sanguine, soit 10,1 mmol/l, sont aussi anormalement élevées. Les autres indices révélés lors de l'anamnèse et de l'examen physique ainsi que par les tests de laboratoire, notamment l'analyse d'urine, l'amènent à poser un diagnostic d'insuffisance rénale aiguë. C'est une raison suffisante pour justifier l'hospitalisation de Sarah¹. Lisa est intriguée par le cas de la jeune femme. En effet, elle n'était pas au courant du lien qu'il pouvait y avoir entre l'ingestion de champignons hallucinogènes et l'insuffisance rénale. Connaît-on une association entre la consommation de tels champignons et l'insuffisance rénale aiguë? S'agit-il d'un cas inusité qui mériterait de faire l'objet d'une publication?

Pertinence clinique

Les connaissances actuelles tirent en bonne partie leur origine d'observations scientifiques faites

par des praticiens ou des chercheurs isolés, observations qui peuvent certes être transmises de vive voix à des pairs. Cependant, afin qu'un plus grand nombre de personnes puissent en bénéficier, elles ont avantagé à être communiquées sous forme d'écrits dans des revues scientifiques. En effet, lorsqu'un médecin ou un chercheur recueille une observation qui lui semble inusitée ou inédite, comme celle dont Sarah a fait l'objet, il doit se demander si le phénomène observé n'est nouveau que pour lui ou si l'ensemble de la communauté scientifique le méconnaît encore à ce jour. En premier lieu, une recherche documentaire et une consultation auprès de ses confrères lui permettront de répondre à son questionnement. En second lieu, si après vérification l'observation s'avérait inédite, il serait pertinent et souhaitable d'en faire part à la communauté scientifique.

Cependant, toute observation inusitée doit-elle faire l'objet d'une publication scientifique? Comment distinguer l'anecdote et le fait scientifique pertinent? À ce chapitre, Lisa aura à prendre une décision en ce qui concerne le cas de Sarah. Elle devra procéder à l'analyse de son observation en fonction de critères de pertinence, comme la survenue d'un fait dans des circonstances inhabituelles, une présentation de signes et de symptômes inusités, des complications méconnues ou un traitement inédit. Le cas échéant, la rédaction d'un article devra suivre quelques règles bien établies.

4.1 L'étude d'un cas

4.1.1 Les caractéristiques de l'étude d'un cas

L'**étude d'un cas** (le *case report* anglo-saxon) est le type d'étude le plus simple et le plus spontané*. Sa conception et sa rédaction sobres en font l'une des formes les plus accessibles et les plus fréquentes de transmission écrite de la connaissance scientifique. Cette façon de communiquer serait d'ailleurs des plus anciennes, car on rapporte qu'Hippocrate l'a utilisée pour ses descriptions écrites de cas cliniques.

Les auteurs d'articles scientifiques choisissent généralement un type d'étude en fonction de la disponibilité des données et de la faisabilité de l'étude. Ainsi, Lisa a reconnu ce qui lui semblait, à première vue, constituer un fait inusité dans sa pratique clinique. Intuitivement, elle a émis l'hypothèse voulant qu'il y ait un lien entre l'ingestion des champignons que Sarah croyait être des champignons «magiques» (à noter qu'il pouvait s'agir d'un autre type de champignons reconnus pour leur toxicité rénale) et l'insuffisance rénale aiguë dont la jeune patiente a souffert.

L'étude d'un cas est souvent le point de départ de la description d'une maladie ou d'un syndrome nouveau ou jusqu'alors méconnu, de signes et de symptômes cliniques inhabituels ou de réactions indésirables importantes à certains médicaments et inconnues à ce jour. Elle peut aussi traiter d'une nouvelle approche diagnostique ou thérapeutique d'un problème de santé notoire, de la manifestation d'une maladie connue observée pour la première fois dans une population particulière, d'un nouveau traitement pour une maladie jusque-là fatale ou d'une relation possible et jusqu'alors insoupçonnée entre deux maladies.

L'étude d'un cas peut aussi être utilisée dans le cadre de l'enseignement aux étudiants ou du développement professionnel continu. On applique alors à un cas complexe une recherche approfondie de données fondées sur des preuves afin d'en élucider le diagnostic ou le choix de traitement². Cet apprentissage peut inciter à se tenir à l'affût du phénomène décrit chez d'autres patients.

L'étude d'un cas permet à l'occasion de suggérer des hypothèses de lien causal entre un facteur de risque et une maladie, mais elle ne permet pas de les vérifier; elles devront être testées au moyen de devis d'étude appropriés. En outre, elle s'avère parfois utile pour décrire un nouveau concept de pathogenèse, pour remettre en question ou parfaire les connaissances existantes ou encore pour stimuler la recherche.

* Il importe de ne pas confondre l'étude «d'un» cas et l'étude «de» cas: en effet, l'étude de cas (le *case study* anglo-saxon) fait appel à une stratégie de recherche qualitative en profondeur, donc différente et plus complexe.

EXEMPLE 4.1

Un médecin exerçant auprès de personnes atteintes de troubles mentaux publie l'étude du cas d'un patient à qui l'on administre du lithium. La concentration plasmatique de ce médicament augmente dramatiquement chez ce patient après un traitement concomitant au kétorolac, un agent anti-inflammatoire non stéroïdien (AINS). Compte tenu de l'interaction médicamenteuse connue des différents AINS avec le lithium, l'auteur propose que l'on ajoute le kétorolac à la liste des produits non recommandés pour les patients traités au lithium³.

4.1.2 Les forces, les faiblesses et les limites de l'étude d'un cas

Malgré sa pertinence et son utilité, l'étude d'un cas présente plusieurs faiblesses du point de vue méthodologique et comporte des limites quant à son interprétation. Par définition, elle porte sur un seul cas sans comparaison avec d'autres cas semblables ou des témoins. Elle ne permet habituellement pas de généraliser et encore moins de confirmer l'hypothèse d'une association entre une exposition et une maladie. Par conséquent, elle ne peut que très exceptionnellement modifier les critères de la prise de décision clinique jusqu'alors acceptés.

4.1.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'un cas (les exigences de publication)

Bien que la facture de l'étude d'un cas puisse varier, elle doit présenter certaines caractéristiques de structure et de contenu qui en attestent la qualité. Cette structure est adaptée à la description d'un cas et diffère ainsi sensiblement de la structure des autres types d'articles scientifiques. En 2013, un consensus d'experts (CARE) a publié un modèle de rédaction d'une étude d'un cas qui a fait l'objet d'une mise à jour en 2016⁴. Nous reprendrons ici l'essentiel de ce consensus.

De façon générale, ce type de publication ne doit pas permettre d'identifier le patient. Les auteurs doivent avoir obtenu, si nécessaire, l'autorisation d'un comité d'éthique. Ils doivent indiquer la présence ou non de conflits d'intérêts. Il importe aussi que les auteurs s'assurent que le patient a donné son consentement éclairé à la publication.

- Le titre : il doit indiquer qu'il s'agit d'une étude d'un cas et ce qui fait l'intérêt de celui-ci. Des mots-clés peuvent aussi désigner les champs couverts par cette étude d'un cas.
- Le résumé : il doit présenter l'intérêt de l'étude du cas. On doit y retrouver succinctement le motif de consultation, les principaux diagnostics, la prise en charge réalisée, l'issue des soins et l'essentiel des enseignements à tirer du cas. Le résumé permet ainsi au lecteur de déterminer rapidement si l'étude est pertinente à sa pratique.
- L'introduction : elle doit mettre en lumière les caractéristiques particulières du cas observé justifiant, par leur caractère unique ou inusité, qu'il fasse l'objet d'une publication au moyen d'explications adéquates, et s'appuyer sur des observations déjà présentées dans la littérature. On y trouvera ainsi les éléments de contexte de l'étude d'un cas.

EXEMPLE 4.2

Lisa consulte quelques collègues omnipraticiens et spécialistes. Elle relève dans la littérature que des cas d'ingestion de champignons toxiques appartenant à certaines espèces d'*Amanita* ou de *Cortinarius* ayant entraîné une insuffisance rénale aiguë ont été signalés en Europe. Bien que ces champignons poussent également en Amérique du Nord, Lisa ne repère aucun rapport de cas semblables survenus sur ce continent. Après avoir obtenu le consentement de Sarah, en la rassurant sur le respect des renseignements personnels ou confidentiels, elle décide donc de soumettre un article portant sur le cas. Les éléments mentionnés ci-dessus feront justement partie de l'introduction de cet article.

La description du cas et le motif de consultation

La description du cas doit être claire et brève. On y trouve les caractéristiques du cas, par exemple l'âge, le sexe, l'origine, l'occupation de la personne (travailleur, étudiant, etc.), ainsi que la présentation clinique ayant motivé la consultation et, s'il y a lieu, les détails pertinents de la prise en charge antérieure.

L'anamnèse et l'examen clinique

On retrouve ici les éléments d'informations pertinentes concernant les antécédents familiaux, médicaux, génétiques et psychosociaux ainsi que celles relatives aux habitudes de vie de la personne. On y décrit aussi les autres problèmes de santé actifs ainsi que leur prise en charge, les traitements reçus de même que l'automédication. Enfin, on y reporte les éléments importants et les signes trouvés lors de l'examen physique.

La dimension chronologique

La chronologie des événements et de l'évolution du cas doit être présentée de façon précise. Ces informations gagnent à être consignées sous la forme d'un tableau ou d'une figure. Le groupe d'experts CARE fournit plusieurs exemples de présentation des éléments chronologiques dans une étude d'un cas⁵.

Le raisonnement et la prise en charge diagnostiques et thérapeutiques

On doit trouver dans l'étude d'un cas les informations pertinentes sur les résultats diagnostiques. Les obstacles diagnostiques seront aussi présentés. Ces obstacles peuvent être par exemple l'impossibilité de mener une évaluation complète compte tenu notamment de barrières culturelles ou d'accès à la technologie. La nature (médicamenteuse, préventive, etc.) et la modalité d'administration (dosage, fréquence, durée, etc.) des interventions réalisées lors de la prise en charge thérapeutique doivent aussi être décrites.

Le suivi et l'issue des soins

On présente ici l'évolution clinique du cas. On y décrit, s'il y a lieu, le maintien, les modifications, les suspensions ou les abandons de traitement et leur justification de même que l'adhésion au traitement et la façon dont elle a été évaluée. Les effets

indésirables ou inattendus des traitements doivent aussi être décrits. Enfin, le résultat des soins selon le patient, l'évaluation clinique du médecin, ainsi que les résultats positifs et négatifs des examens complémentaires importants devraient en faire partie.

EXEMPLE 4.3

Dans son article, Lisa décrit bien la suite des événements qui ont précédé l'apparition des symptômes chez sa patiente selon la séquence temporelle, les symptômes eux-mêmes, l'absence des effets hallucinogènes attendus, les résultats anormaux de l'analyse d'urine, l'évolution vers la guérison spontanée de l'insuffisance rénale et l'absence d'autres facteurs connus qui auraient pu entraîner cette affection. Son raisonnement lui permet de constater que tout concorde pour conclure à l'ingestion de champignons de type *Cortinarius* comme cause probable de la maladie plutôt que l'ingestion de ce que Sarah croyait être des champignons « magiques ».

La description que Lisa fait du cas permet d'en arriver au diagnostic d'insuffisance rénale aiguë d'origine toxique. Elle met en lumière le lien temporel entre l'ingestion de ce champignon non hallucinogène connu pour entraîner cette affection et l'apparition subséquente de la maladie.

La discussion

Cette section de l'article compare l'étude en question avec d'autres études semblables, le cas échéant, et décrit l'étendue et l'ampleur de la recherche documentaire, si cela n'a pas déjà été fait dans l'introduction. Les auteurs doivent aussi discuter des questions suivantes : « Les observations cliniques présentées justifient-elles le diagnostic posé ? », « Y a-t-il d'autres explications plausibles ? » et « Comment interprète-t-on les éventuelles observations contradictoires ? » En dernier lieu, la discussion présente les implications de l'étude pouvant conduire le lecteur à modifier ses vues sur la question, ainsi que des indications sur la direction que pourraient prendre des études futures ou sur le traitement de cas semblables dans la mesure où d'autres études plus poussées confirmeraient subséquentement les hypothèses qui y sont posées.

En somme, la discussion permet de décrire les points forts et les limites de l'étude d'un cas présenté en référence à la littérature en sciences de la santé. Elle donne aussi l'occasion de justifier les conclusions des auteurs et de mettre en évidence l'essentiel de ce qu'il faut retenir du cas.

EXEMPLE 4.4

Lisa discute des mécanismes physiopathologiques connus de la toxicité grâce à des études menées sur des animaux. Enfin, elle termine son article en mentionnant que, compte tenu de l'engouement pour les champignons hallucinogènes et des erreurs fréquentes dans l'identification des différentes espèces de champignons, on devra envisager, dans les cas d'insuffisance rénale aiguë inexplicée, qu'une intoxication due à l'ingestion de champignons puisse être en cause.

4.2 L'étude d'une série de cas

4.2.1 Les caractéristiques de l'étude d'une série de cas

L'étude d'une série de cas est une extension de l'étude d'un cas. Outre les applications décrites précédemment pour l'étude d'un cas, elle peut aussi permettre de décrire les circonstances d'apparition de maladies rares et de donner des indications sur les caractéristiques générales des personnes atteintes (âge, sexe, etc.), ainsi que sur les facteurs associés à la maladie, sur le pronostic, le traitement et la prévention. La réalisation d'une étude de ce type peut également rendre compte d'éléments utiles à la pratique, comme une description des répercussions de la maladie sur la famille et sur l'utilisation des ressources communautaires ou un portrait plus détaillé de l'évolution de la maladie au fil du temps. Ainsi, l'étude d'une série de cas sera utile pour faire apparaître des éléments de connaissance permettant de formuler de nouvelles hypothèses, de rapporter des événements sentinelles concernant par exemple les effets indésirables de certaines thérapies, la survenue d'épidémies ou l'apparition de syndromes nouveaux ou méconnus. Elle pourra aussi s'avérer utile, mais de façon habituellement plus limitée, pour l'étude de maladies rares ou de nouveaux traitements⁶.

EXEMPLE 4.5

Pendant l'été 1976, une épidémie de pneumonie s'est déclarée à Philadelphie lors d'une convention de la Légion américaine à laquelle participaient environ 5 000 personnes. L'épidémie a touché 216 personnes et causé 29 décès. La description de cette série de cas a fait ressortir notamment une caractéristique commune : ces personnes avaient toutes fréquenté le lobby d'un même hôtel avant que la maladie se déclare. Aucun microorganisme n'avait été mis en évidence à ce moment-là. Ce n'est que plus tard que les cultures et les examens histopathologiques des lésions pulmonaires ont permis d'isoler une bactérie jusqu'alors inconnue qui se serait dispersée à partir du système de climatisation de l'hôtel. On a nommé cette bactérie *Legionella pneumophila* et donné le nom de « maladie du légionnaire » ou « légionellose » à la nouvelle maladie, en mémoire des victimes de l'épidémie de Philadelphie. Des études rétrospectives ont par la suite permis de déceler des cas méconnus de légionellose datant de 1943. Plusieurs études et investigations d'épidémie ont subéquemment indiqué que les bactéries du genre *Legionella* sont une cause courante de pneumonie dans la population et demeurent souvent à l'origine d'infections contractées à l'hôpital (infections nosocomiales). L'évolution des connaissances a permis d'établir que les chauffe-eau et les tours de refroidissement à l'eau constituaient deux des sources possibles de contamination.

EXEMPLE 4.6

Le 5 juin 1981, la revue américaine *Morbidity and Mortality Weekly Report*⁷ publiait une étude d'une série de cinq cas de patients atteints d'une pneumonie à *Pneumocystis carinii*; les cinq patients étaient hospitalisés dans trois hôpitaux de Los Angeles, en Californie. Cette observation semblait alors inusitée, car cette pneumonie, considérée comme une infection opportuniste très rare à l'époque,

EXEMPLE 4.6 (suite)

ne touchait que des personnes souffrant d'une immunodéficience marquée. Les caractéristiques rapportées dans l'étude mentionnaient que les cinq patients ne se connaissaient pas et n'avaient eu aucun contact commun. Cependant, tous étaient de sexe masculin et homosexuels.

Un mois plus tard, la même revue rapportait 26 cas de sarcome de Kaposi, un cancer rare aux États-Unis, survenu aussi chez des hommes homosexuels habitant New York et la Californie. Une année plus tard, on rapportait, chez des personnes auparavant en bonne santé, 355 cas de ce que l'on désignait toujours à l'époque par l'appellation hybride « sarcome de Kaposi et infections opportunistes », ces dernières incluant la pneumonie à *P. carinii*.

Finalement, c'est en septembre 1982 que l'on a commencé à utiliser le vocable maintenant bien connu de « syndrome de l'immunodéficience humaine acquise » ou « sida ». À ce moment déjà, l'analyse des caractéristiques épidémiologiques des cas rapportés laisse soupçonner une cause infectieuse, bien que le virus ne soit pas encore isolé. Parmi les facteurs de risque alors identifiés figuraient l'homosexualité ou la bisexualité, l'utilisation de drogues intraveineuses et l'hémophilie A. Les hypothèses de transmission par contact sexuel ou par contact avec le sang ou ses sous-produits sont toutes issues de ces études d'un cas et de séries de cas. Parmi les facteurs de risque, on a ensuite retranché l'orientation sexuelle et plutôt fait référence aux comportements sexuels à risque.

EXEMPLE 4.7

Pendant l'été 1999, l'observation de deux cas inexplicables de fièvre et d'encéphalite a été rapportée aux autorités de santé publique par un infectiologue astucieux. Au même moment, on avait remarqué le décès de corneilles sauvages et d'oiseaux du zoo du Bronx avec évidence d'hémorragie cérébrale. L'enquête et les recherches en laboratoire ont permis d'identifier le virus du Nil occidental comme étant la cause de ces événements⁸. Ce virus était jusqu'alors absent du continent américain. La confirmation de l'introduction de ce nouveau virus avec son potentiel épidémique a entraîné des campagnes préventives importantes.

EXEMPLE 4.8

Dès février 2016, le Comité d'urgence du Règlement sanitaire international concernant le virus Zika concluait que l'agrégat (regroupement) récent de cas de microcéphalie et d'autres troubles neurologiques signalés au Brésil, faisant suite à un regroupement similaire de cas en Polynésie française en 2014, constituait une urgence de santé publique de portée internationale (USPPI)⁹. Ce constat a été jugé suffisant pour recommander la mise en œuvre de messages et d'actions préventives à portée mondiale.

À la limite, lorsqu'une maladie n'a effectivement qu'une cause unique, la découverte de cette dernière permet de déduire un lien de cause à effet entre celle-ci et la maladie.

EXEMPLE 4.8 (suite)

Par ailleurs, dans certaines situations, comme l'apparition de quelques cas d'une maladie très rare, l'étude d'une série de cas permet de formuler des hypothèses causales très justes.

EXEMPLE 4.9

Actuellement, plusieurs travailleurs en usine bénéficient de certaines mesures de protection et de surveillance médicale. Le cas particulier de la survenue du cancer du foie chez quelques travailleurs exposés au chlorure de vinyle offre un exemple de l'importante contribution des études de série de cas. En effet, en 1974, deux médecins américains rapportaient trois cas d'angiosarcome du foie, un cancer très rare de cet organe, chez des travailleurs d'une usine de polymérisation de chlorure de vinyle¹⁰. Cette observation a fourni une preuve raisonnable qui a permis de considérer l'exposition au chlorure de vinyle comme un facteur déterminant dans les cas observés. Des études épidémiologiques plus poussées ont par la suite confirmé cette association.

4.2.2 Les forces, les faiblesses et les limites de l'étude d'une série de cas

L'étude d'une série de cas présente, à quelques variantes près, les mêmes faiblesses méthodologiques et limites que celles associées à l'étude d'un cas quant à son interprétation. Ses faiblesses tiennent au fait qu'elle porte en général sur un petit nombre de patients souvent sélectionnés en fonction de la disponibilité des données et non de la représentativité de la population à laquelle ils appartiennent. Par ailleurs, il n'y a pas, dans ce type d'étude, de sujets témoins ou d'autres cas auxquels on pourrait comparer les patients décrits dans la série de cas. Ainsi, on ne peut presque jamais, en se basant sur de telles études, démontrer des associations entre un facteur de risque et une maladie. On ne peut guère davantage s'en servir pour prouver l'efficacité d'un traitement. Contrairement aux autres types d'études, les conclusions des études d'une série de cas ne peuvent habituellement pas être généralisées à toute la population. Par conséquent, elles ne peuvent que très exceptionnellement modifier les critères de la prise de décision clinique jusqu'alors acceptés.

4.2.3 L'évaluation de la qualité des publications scientifiques portant sur l'étude d'une série de cas (les exigences de publication)

Les critères généraux de publication d'un article scientifique s'appliquent plus aisément à une étude d'une série de cas que lorsqu'il s'agit d'une étude d'un cas.

- Le titre et le résumé : le titre doit être évocateur de ce genre d'études. Le résumé constitue une partie essentielle de l'étude d'une série de cas. Il s'agira généralement d'un résumé structuré, c'est-à-dire que chacune des parties de l'article y sera brièvement présentée, sans omission.
- L'introduction : elle doit contenir les mêmes éléments que ceux que comporte une étude d'un cas.

- **La méthode :** une section « Méthode » doit décrire clairement le procédé au moyen duquel les cas présentés ont été retenus, soit la définition de cas utilisée, la provenance de ceux-ci, les observations faites et la façon dont celles-ci ont été recueillies. Le lecteur devrait pouvoir ainsi juger si les résultats de l'étude peuvent s'appliquer à ses patients. De fait, une série de cas décrits dans le cadre d'une pratique spécialisée en centre hospitalier de référence de troisième ligne pourrait être fort différente de celle de cas similaires observés dans une pratique de première ligne.
- **Les résultats :** la section « Résultats » doit passer en revue les caractéristiques générales des patients, c'est-à-dire les variables telles que l'âge, le lieu de résidence, la saison et les autres observations spécifiques et pertinentes dans le contexte de l'étude présentée. Non seulement y précisera-t-on les fréquences absolues et relatives, mais on y résumera souvent les informations à l'aide de mesures de tendance centrale, comme le mode, la moyenne et la médiane, et de mesures de dispersion, comme l'étendue et l'écart type. Les résultats les plus intéressants seront présentés à l'aide de tableaux ou de graphiques.
- **La discussion :** cette section doit comporter les mêmes éléments que ceux qui figurent dans l'étude d'un cas.

4.3 Une grille d'analyse des études d'un cas et d'une série de cas

La grille présentée dans le tableau 4.1 constitue une liste de vérification concernant l'information à inclure dans l'étude d'un cas. Elle peut aussi s'avérer utile pour l'évaluation de la qualité de la forme et du contenu d'un article portant sur une série de cas.

TABLEAU 4.1 ■ Une liste de vérifications de l'information à inclure dans les études d'un cas

Section	Élément	Description
Titre	1	Le titre devrait comporter les termes « rapport » ou « étude d'un cas » avec l'élément d'intérêt principal.
Mots clés	2	Ce sont de quatre à sept mots clés précisant les domaines couverts par le rapport de cas.
Résumé	3a	Introduction : qu'est-ce qui rend ce ou ces cas uniques ? Qu'apportent-ils à la littérature médicale ?
	3b	Résumé du cas : principaux symptômes et signes cliniques importants, principaux diagnostics, traitements et issue des soins.
	3c	Conclusion : quels sont les principaux enseignements à tirer de ce cas ?
Introduction	4	Elle se compose d'un ou de deux paragraphes qui résument pourquoi ce cas est unique, avec des références à l'appui.
Chronologie	5	Elle relate les moments importants de l'histoire du patient, présentés de façon chronologique (tableau ou figure).
Informations sur le patient	6a	Ce sont des informations démographiques et d'autres caractéristiques du patient qui ne peuvent toutefois mener à son identification. →

TABLEAU 4.1 ■ Une liste de vérifications de l'information à inclure dans les études d'un cas (suite)

Section	Élément	Description
Informations sur le patient	6b	Elles précisent les plaintes et les symptômes principaux du patient ayant motivé la consultation.
	6c	Elles comprennent les antécédents familiaux, médicaux et psychosociaux en incluant les informations génétiques pertinentes, la prise en charge antérieure avec son résultat.
Examen clinique	7	Il faut y décrire les éléments pertinents de l'examen dirigé et les anomalies importantes de l'examen complet.
Évaluation diagnostique	8a	Elle comprend les examens réalisés, notamment l'examen physique, les analyses de laboratoire, l'imagerie, les questionnaires.
	8b	Elle englobe les raisonnements diagnostiques avec les hypothèses évoquées, les autres diagnostics considérés, les défis ou les difficultés rencontrés.
	8c	Il faut considérer l'utilisation de tableaux ou de figures pour faire le lien entre l'évaluation, le diagnostic et l'intervention.
	8d	Elle comporte les éléments pronostiques, le cas échéant.
Intervention	9a	Elle décrit les types d'interventions réalisées (recommandations sur les habitudes de vie, médicamenteuses, chirurgicales, préventives, soins autoadministrés).
	9b	Elle précise les modalités d'administration des traitements telles que la fréquence, le dosage, la durée.
	9c	Elle détaille les modifications de la prise en charge thérapeutique prévue avec leur justification.
	9d	Elle précise les autres interventions concomitantes.
Suivi et issue des soins	10a	On y indique l'issue des soins du point de vue du médecin et du patient (si approprié).
	10b	Les évaluations des suivis diagnostiques importants y sont consignées.
	10c	L'évaluation de l'adhésion au traitement et de sa tolérance y est indiquée, incluant les effets indésirables ou inattendus.
Discussion	11a	Elle souligne les points forts et les limites de la prise en charge du cas.
	11b	Elle spécifie en quoi l'étude d'un cas contribue à la pratique et aux guides de pratique clinique.
	11c	Elle pose les hypothèses vérifiables que suggère l'étude d'un cas.
	11d	Elle précise les messages clés, les conclusions et la pertinence de ce rapport de cas.
Point de vue du patient	12	Quand cela est approprié, le patient devrait relater son expérience des soins reçus.
Consentement éclairé	13	Le patient a-t-il donné un consentement éclairé? (À fournir sur demande)
Information additionnelle	14	Ce sont les sections sur les remerciements, les conflits d'intérêts, l'approbation par un comité d'éthique.

En complément à cette liste, pour l'évaluation de la qualité, on devra trouver dans une série de cas une description précise des mesures de fréquence utilisées. L'encadré 4.1 résume les forces et les faiblesses des études d'un cas et d'une série de cas.

ENCADRÉ 4.1 ■ Les forces et les faiblesses des études d'un cas et d'une série de cas

Forces communes aux deux types d'études

- Permettent la détection rapide et la description de :
 - maladies et syndromes nouveaux ou jusqu'alors méconnus ;
 - présentations cliniques inattendues ou inhabituelles d'une maladie ;
 - réactions indésirables ou bénéfiques importantes aux médicaments, ou d'interactions médicamenteuses jusqu'alors inconnues ;
 - nouvelles relations ou variations dans le développement des maladies ;
 - nouvelles approches diagnostiques ou thérapeutiques ;
 - la manifestation d'une maladie connue dans une population jusqu'alors non touchée ;
 - nouveaux traitements pour des maladies jusque-là fatales ;
 - relations possibles et jusqu'alors insoupçonnées entre deux maladies ;
 - l'évolution de la maladie au fil du temps, de manière plus détaillée ou inattendue.
- Peuvent mener à des hypothèses causales entre une exposition et une maladie.
- Complètent ou remettent en question des connaissances existantes sur les maladies, leurs causes et leur traitement.
- Stimulent la recherche et servent pour l'enseignement.

Forces exclusives aux études d'une série de cas

- Permettent la détection rapide et la description des :
 - mécanismes de transmission des maladies ;
 - répercussions de la maladie sur la famille et la communauté ;
 - répercussions de la maladie sur l'utilisation des ressources communautaires.
- Donnent des indications sur la fréquence des facteurs associés à la maladie, sur le pronostic, le traitement et la prévention.
- Peuvent être utilisées pour l'évaluation de la qualité de l'acte (assurance qualité).

Faiblesses communes aux deux types d'études

- Ne permettent généralement pas de :
 - généraliser les résultats de l'observation à toute la population ;
 - confirmer une hypothèse d'association entre un facteur d'exposition et une maladie.
 - servir de preuve de l'efficacité d'un traitement ;
 - modifier les critères de la prise de décision clinique jusqu'alors acceptés, sauf très exceptionnellement dans le cas de l'étude d'une série de cas.

Faiblesses exclusives aux études d'une série de cas

- Se fondent généralement sur un petit nombre de cas sélectionnés en fonction de la disponibilité des données.
- Ne comparent pas les cas choisis à des sujets témoins.

4.4 Les enjeux éthiques pour les études d'un cas ou d'une série de cas

Le professionnel de la santé a le devoir de partager ses connaissances avec ses pairs en faisant preuve de rigueur scientifique et de prudence et en prenant en considération les connaissances les plus à jour.

Il doit aussi tenir compte de l'autonomie du patient qui pourrait refuser que l'étude de son cas soit publiée à cause de risques possibles de se faire identifier par des tiers, voire d'être stigmatisé. Il est important d'obtenir le consentement éclairé du patient. À défaut, il faut obtenir l'autorisation des autorités compétentes.

Enfin, il doit assurer la protection et la confidentialité des renseignements personnels et cliniques publiés, en n'incluant que ceux qui sont nécessaires.

En résumé

L'étude d'un cas est une forme de communication fréquemment rencontrée dans la littérature scientifique. Lisa l'a utilisée pour faire part à la communauté scientifique de son observation d'un cas d'insuffisance rénale aiguë causée par une intoxication aux champignons dont il est permis de croire qu'ils sont de type *Cortinarius*. Cette étude d'un cas survenu au Canada attire l'attention du lecteur sur la possibilité que l'on trouve sur le continent nord-américain une telle cause d'insuffisance rénale, jusqu'alors rapportée uniquement en Europe.

L'étude d'un cas peut également élargir les horizons et favoriser l'acquisition de nouvelles connaissances au sujet d'une maladie, de son évolution, de ses conséquences ou de son traitement. Certains critères permettent d'en apprécier la qualité, tant sur le plan de la forme que sur celui du contenu. L'étude d'un cas peut mener à des hypothèses, mais ne les confirme pas; elle informe le chercheur sur une situation particulière, mais ne permet pas de la généraliser à toutes les situations semblables; et elle complète ou remet en question

des connaissances déjà existantes. L'étude d'une série de cas possède toutes les forces et les limites de l'étude d'un cas. En outre, elle permet au lecteur de prendre connaissance des mesures de fréquence des caractéristiques des cas ainsi que des observations spécifiques et pertinentes dans le contexte particulier de l'étude. Plus rarement, elle peut suggérer un lien de causalité entre une maladie et une exposition à un ou plusieurs facteurs, et appuyer des actions préventives.

À une époque où l'apprentissage repose sur les données fondées sur des preuves, l'étude d'un cas et l'étude d'une série de cas constituent des formes de communication des connaissances ayant une grande capacité de détecter ou de découvrir des faits jusque-là inconnus ou non reconnus. Les études à visée étiologique viendront ensuite confirmer ou infirmer ces faits.

Somme toute, l'étude d'un cas et l'étude d'une série de cas ont pleinement leur place en épidémiologie. Elles représentent aussi un moyen privilégié d'enseignement et d'assurance qualité.

Exercices**Voir la section complète des exercices sur le Web.**

1. Vrai ou faux. Une étude d'une série de cas permet de suggérer des hypothèses causales sans toutefois permettre de les vérifier.
2. Choisissez la bonne réponse pour compléter la phrase suivante: La principale justification de la publication d'une étude d'un cas réside dans le fait:
 - a) que l'auteur s'appuie sur un grand nombre d'années d'expérience dans le domaine de la publication.
 - b) qu'il s'agit d'un cas très rare.
 - c) que les observations contribuent à l'avancement des connaissances.
3. Vrai ou faux. Dans une étude d'un cas, la population d'où le cas est issu doit être décrite adéquatement.
4. Vrai ou faux. L'étude d'un cas permet d'offrir à des patients un traitement semblable à celui dont il est question dans l'étude.
5. Vrai ou faux. Les études d'un cas ou d'une série de cas constituent des moyens de communication des connaissances susceptibles de mettre en évidence des faits inconnus ou non reconnus.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 5

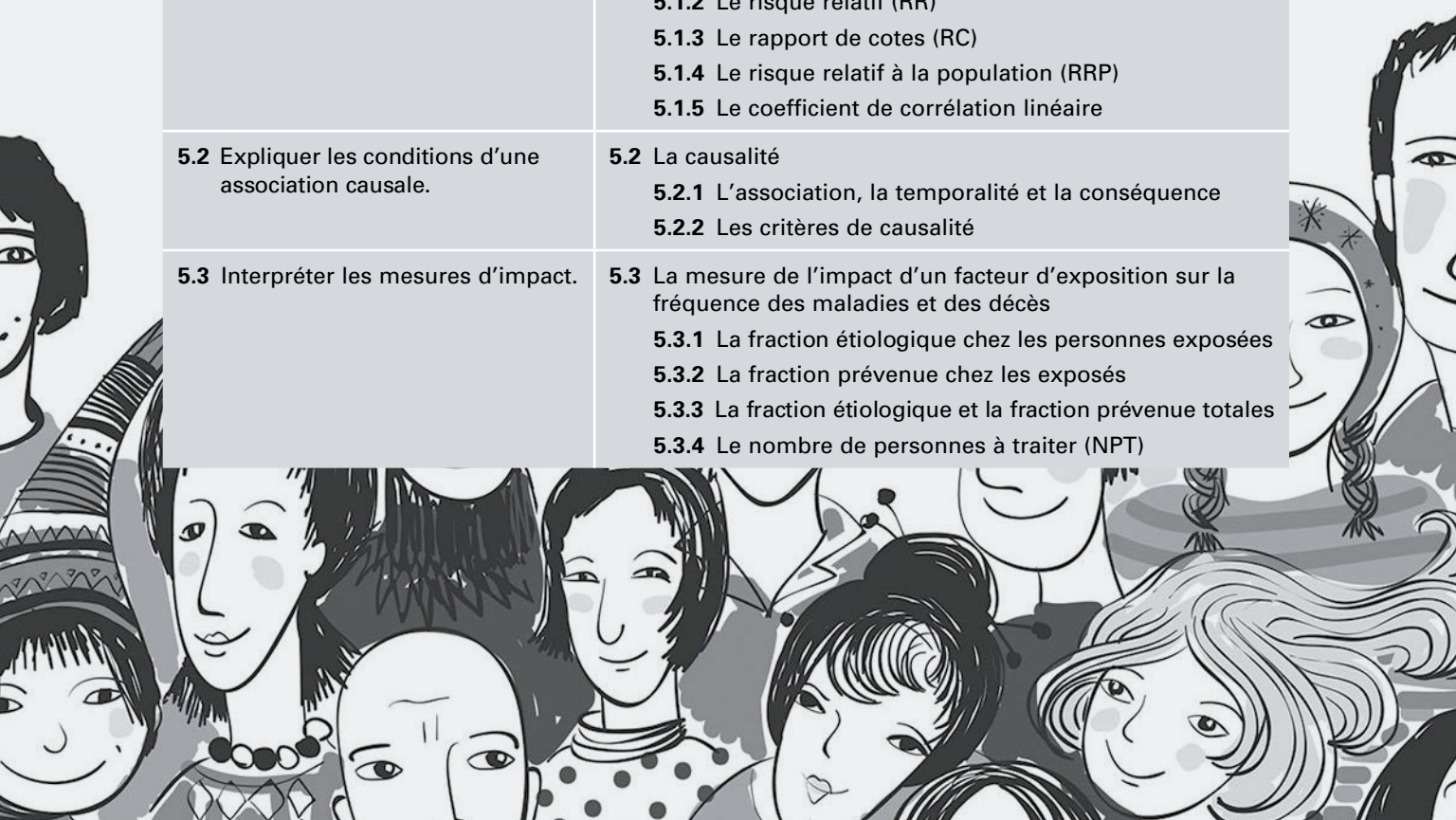
Les mesures d'association et la causalité

Collaboration à la 3^e édition :
André Simpson
Thomas Matukala Nkosi

Au terme de ce chapitre, vous pourrez :

- expliquer les concepts et les méthodes en épidémiologie qui permettent de déterminer les causes des maladies et des décès ;
- interpréter les mesures en épidémiologie qui permettent de vérifier la présence et de quantifier la force et l'impact d'une association entre un facteur de risque et une maladie ou les décès ;
- vérifier si une association observée permet de conclure à une relation de cause à effet entre un facteur de risque et la maladie ou les décès.

Objectifs spécifiques	Contenu
5.1 Interpréter les mesures d'association.	5.1 Les mesures d'association 5.1.1 Le risque attribuable (RA) 5.1.2 Le risque relatif (RR) 5.1.3 Le rapport de cotes (RC) 5.1.4 Le risque relatif à la population (RRP) 5.1.5 Le coefficient de corrélation linéaire
5.2 Expliquer les conditions d'une association causale.	5.2 La causalité 5.2.1 L'association, la temporalité et la conséquence 5.2.2 Les critères de causalité
5.3 Interpréter les mesures d'impact.	5.3 La mesure de l'impact d'un facteur d'exposition sur la fréquence des maladies et des décès 5.3.1 La fraction étiologique chez les personnes exposées 5.3.2 La fraction prévenue chez les exposés 5.3.3 La fraction étiologique et la fraction prévenue totales 5.3.4 Le nombre de personnes à traiter (NPT)





Mise en situation

Il y a trois ans, au cours d'une visite en milieu scolaire, Danielle, infirmière à la polyclinique Milo, reçoit Rabi, un solide gaillard de 15 ans qui la consulte à propos d'une acné qui le gêne beaucoup.

« J'ai entendu dire que les filles de l'école se préparaient une tisane à partir d'une plante (elles appellent ça du "petit thé"), et que cela faisait disparaître leur acné. Est-ce que ça pourrait régler mon problème ? »

Danielle ne connaît pas le « petit thé » en question. Elle décide de se renseigner. Le « petit thé » serait en fait une infusion de chiogène hispide, un arbuste nordique, rampant, à petites feuilles alternes, dont l'arôme rappelle l'essence de wintergreen. La littérature officielle est muette quant à ses vertus thérapeutiques.

Puisque l'usage de cette infusion est assez répandu chez les adolescentes de la région,

l'équipe propose de mener une étude pour déterminer s'il existe vraiment une association entre ladite infusion et l'acné et, le cas échéant, pour savoir si l'on peut établir un lien de cause à effet entre la consommation d'une infusion de chiogène hispide et la diminution de l'acné.

Pertinence clinique

Un bon nombre de traitements traditionnels s'avèrent farfelus, mais plusieurs ont une efficacité réelle. Comment peut-on les départager ? L'épidémiologie offre des modèles et des types d'études permettant de cerner cette réalité. En épidémiologie, il existe une série de mesures permettant de dégager des associations ou des liens entre les variables catégorielles ou quantitatives observées, pour quantifier l'intensité et l'impact de ces associations.

Le fait de mettre en lumière une association entre deux variables implique-t-il obligatoirement que l'une d'elles soit la cause de l'autre ? Pas nécessairement. Les critères de causalité permettent de guider le jugement à cet égard.

L'équipe de la polyclinique a donc décidé de mener une étude à visée étiologique de type expérimental au cours de laquelle elle a comparé deux cohortes d'adolescentes. Danielle a dirigé l'étude.

Désignés par le hasard, un premier groupe a consommé le « petit thé » (groupe T+); et un second, un placebo (groupe T-). Marie-Louise, la pharmacienne à la polyclinique, s'est chargée de préparer les infusions; elle était la seule à pouvoir reconnaître la substance à l'étude. Il s'agissait donc d'un essai thérapeutique en double aveugle, ou à double insu. Puisque l'acné évolue par épisodes intermittents, la durée des épisodes a été notée pour les deux groupes.

Pendant les trois années qu'a duré l'étude, tout s'est bien déroulé. Toutes les adolescentes ont été suivies du début à la fin; aucune n'a été perdue de vue. On a ainsi pu mesurer l'incidence cumulée et le taux d'incidence des épisodes d'acné dans les deux groupes pendant toute la période.

Certains résultats sont présentés dans les tableaux 5.1 et 5.2 (à noter que les données présentées dans ces tableaux sont fictives et fantaisistes).

TABLEAU 5.1 ■ L'acné chez des adolescentes de 11 ans à 15 ans consommant du « petit thé » ou un placebo et suivies pendant 3 ans

	Groupe T+ (« petit thé »)	Groupe T- (placebo)	Total
A fait de l'acné	55	65	120
N'a pas fait d'acné	45	35	80
Total	100	100	200

TABLEAU 5.2 ■ Le nombre d'épisodes d'acné chez des adolescentes de 11 ans à 15 ans consommant du « petit thé » ou un placebo et suivies pendant 3 ans

	Groupe T+	Groupe T-	Total
Nombre d'épisodes d'acné	138	168	306
Expérience en années*	275	285	560

* Cumul des années sans acné (années à risque) dans les deux groupes.

5.1 Les mesures d'association

Les **mesures d'association** permettent de vérifier s'il existe une relation entre un ou plusieurs facteurs et la présence ou l'absence de la maladie ou des décès. Elles permettent également de mesurer la force de cette association.

5.1.1 Le risque attribuable (RA)

Le **risque attribuable (RA)** permet de mesurer la différence entre les risques d'être atteint de la maladie ou de décéder chez les personnes exposées et chez les personnes non exposées. On soustrait à cette fin la fréquence de la maladie ou des décès observée chez les personnes non exposées à la même mesure de fréquence chez les personnes exposées. Le risque attribuable est utilisé dans les études à visée étiologique de cohorte parce qu'il compare la fréquence de la maladie ou des décès chez des individus exposés à un facteur de risque à la fréquence de la maladie ou des décès chez des individus non exposés.

En consultant le tableau 5.1, on constate que dans le groupe T+, le risque d'être atteint d'au moins un épisode d'acné est de 55 %, alors qu'il est de 65 % dans le groupe T-. La différence de risque est donc de 55 % – 65 %, soit –10 %. On nomme cette différence le « risque attribuable » (RA). Lorsque le risque est le même dans les deux groupes, le risque attribuable est égal à 0. C'est ce que l'on appelle la « valeur

nulle» pour le RA. (On verra dans le chapitre 6 que la différence observée ici peut être due en partie à l'effet du chiogène hispide, en partie au hasard et en partie à des erreurs qui peuvent fausser les résultats des études. Supposons pour le moment que l'étude ne comporte pas d'erreurs et que le hasard n'ait pas influencé le résultat; la différence de risque est donc attribuable à l'effet du «petit thé».)

À la lumière de cette différence, on peut affirmer que les jeunes filles de 11 ans à 15 ans qui consomment du «petit thé» risquent* un peu moins de faire de l'acné que celles qui n'en consomment pas. C'est donc dire que le chiogène hispide aurait un certain effet protecteur contre l'acné. En effet, ce que le risque attribuable indique, c'est que la différence de 10 % constatée entre les deux groupes est probablement due à la consommation du «petit thé». Autrement dit, pour 100 jeunes filles qui prennent du «petit thé», on préviendrait l'apparition d'acné chez 10 d'entre elles. Donc, lorsque la valeur du risque attribuable est négative, elle révèle un effet protecteur de l'exposition concernée.

Évidemment, le risque attribuable n'est pas toujours négatif, comme le montre l'exemple suivant.

EXEMPLE 5.1

Dans une étude britannique où des médecins de 35 ans et plus, fumeurs et non-fumeurs, ont été observés sur une période de quatre ans et demi¹, on a trouvé que le taux de mortalité par cancer du poumon (ajusté selon l'âge) était de $0,07 \times 10^{-3} \text{ an}^{-1}$ pour les non-fumeurs, alors qu'il s'élevait à $0,9 \times 10^{-3} \text{ an}^{-1}$ pour les fumeurs. La différence de risque (exprimée ici comme une différence de taux) est dans ce cas de $0,83 \times 10^{-3} \text{ an}^{-1}$. On peut donc penser que le fait de fumer entraîne un excès de 83 décès par cancer du poumon par 1 000 fumeurs-année.

En somme, dans le contexte d'une association protectrice, le RA permet d'estimer le nombre de cas que le facteur étudié a réussi à prévenir. En revanche, dans le contexte d'une association causale, le RA permet d'estimer le nombre de cas attribuables au facteur ou, en d'autres mots, le nombre de cas qui auraient pu être évités si l'exposition à ce facteur n'avait pas eu lieu.

5.1.2 Le risque relatif (RR)

Le **risque relatif (RR)** est le rapport entre le risque d'être atteint de la maladie ou de décéder observé chez des personnes exposées au facteur étudié et le risque d'être atteint de la maladie ou de décéder observé chez des personnes non exposées à celui-ci. Ce risque d'être atteint de la maladie peut être mesuré à l'aide d'une mesure de fréquence tels une prévalence, un taux d'incidence, une incidence cumulée ou un taux de mortalité.

Ainsi, on peut également considérer l'association entre deux variables en faisant le rapport de leurs mesures. Lorsque ces mesures sont des proportions ou des taux,

* Il importe de noter que, dans le contexte de la statistique et de l'épidémiologie, les termes «probabilité», «chance» et «risque» sont employés comme des synonymes et ne véhiculent aucune connotation affective. Ainsi, on emploiera indistinctement les expressions «probabilité de mourir», «chance de mourir» ou «risque de mourir», tout comme on parlera indifféremment de la «probabilité», de la «chance» ou du «risque» d'obtenir un effet protecteur pour la santé.

le rapport des mesures s'appelle «risque relatif» (RR)*. Comme le risque attribuable, le risque relatif est utilisé dans les études à visée étiologique de cohorte parce qu'il compare la fréquence de la maladie ou des décès chez des individus exposés au facteur de risque étudié à la fréquence de la maladie ou des décès chez des individus non exposés.

Le rapport de proportions (d'incidences cumulées ou de prévalences)

Pour mesurer la force de l'association entre la consommation de «petit thé» et l'acné, Danielle a calculé le rapport des incidences cumulées. À la lumière des données recueillies (voir le tableau 5.1), le rapport des risques (celui des adolescentes exposées par rapport à celui des adolescentes non exposées), ou risque relatif, est de :

$$\frac{55/100}{65/100}, \text{ soit } 0,85.$$

Lorsque les risques sont les mêmes dans les deux groupes, le rapport des risques est égal à 1. Un tel rapport indique qu'il n'y a pas d'association entre le facteur étudié et l'effet observé. C'est ce que l'on appelle la «valeur nulle» pour le RR.

Lorsque le rapport est différent de 1, c'est qu'il existe une association entre le facteur étudié et l'effet observé. Si le rapport est plus petit que 1, cela signifie que le risque pour les personnes exposées au facteur a été moins grand que pour les personnes non exposées. Le facteur «protège» en quelque sorte les personnes qui y ont été exposées, et l'association est dite négative. En revanche, un rapport plus grand que 1 indique que le risque pour les personnes exposées au facteur a été plus grand que pour les personnes non exposées. Dans ce cas, le facteur augmente la probabilité d'observer l'effet chez les personnes exposées, et l'association est dite positive. Plus la valeur du RR s'éloigne de 1 (valeur nulle), plus forte est l'association observée.

Danielle pourrait par conséquent conclure que le «petit thé» protège contre l'acné les personnes qui en consomment.

Le rapport de taux (d'incidence ou de mortalité)

Lorsqu'une étude a inclus la mesure des taux d'incidence ou de mortalité, il est possible de calculer le risque relatif à partir du rapport de ces taux. Si les taux sont les mêmes dans les deux groupes, le rapport des taux sera égal à 1. Un rapport égal à 1 indique qu'il n'y a pas d'association entre le facteur étudié et l'effet observé.

En s'appuyant sur les données du tableau 5.2, Danielle constate que le taux d'incidence des épisodes d'acné dans le groupe ayant consommé du «petit thé» est de 502 pour 1 000 personnes-année à risque, alors qu'il est de 590 pour 1 000 personnes-année à risque dans le groupe n'ayant pas consommé de «petit thé». Le risque relatif obtenu à partir de ces taux d'incidence est le suivant : $502/590 = 0,85$.

Comme c'est le cas en ce qui concerne le rapport de proportions, lorsque le rapport de taux est différent de 1, cela signifie qu'il existe une association entre le facteur étudié et

* Plusieurs auteurs utilisent aussi une notation propre aux mesures utilisées. Ainsi, on désigne souvent un rapport de prévalences par RP, un rapport d'incidences cumulées par RIC, un rapport de taux d'incidence par RI ou RTI ou RT, un rapport de taux de mortalité par RM ou RTM, etc. Ce sont tous des risques relatifs; dans ce chapitre, on utilisera une notation unique pour les désigner, soit RR.

l'apparition de la maladie ou des décès. Si le rapport est plus petit que 1, cela révèle que le facteur diminue la probabilité que la maladie ou les décès surviennent, et l'association est alors négative. Si le rapport est plus grand que 1, l'association est positive puisque le facteur augmente la probabilité que la maladie ou les décès surviennent. Ici également, plus la valeur du RR est éloignée de 1 (valeur nulle), plus forte est l'association observée. Il importe également de retenir que si, dans une même étude, comme c'est le cas dans celle de Danielle, on calcule le risque relatif à l'aide du rapport d'incidences cumulées (rapport de proportions selon le tableau 5.1) ou à l'aide du rapport de taux d'incidence (rapport de taux selon le tableau 5.2), le résultat du risque relatif sera le même.

EXEMPLE 5.2

Les figures 5.1 et 5.2 illustrent les risques relatifs de mortalité (rapports de mortalité) par cancer du poumon chez les fumeurs, calculés en fonction, respectivement, du nombre de cigarettes qu'ils consomment chaque jour et de l'âge qu'ils avaient lorsqu'ils ont commencé à fumer².

On note à la figure 5.1 que le risque de mourir par cancer du poumon augmente avec le nombre de cigarettes fumées chaque jour puisque la valeur du RR, donc la force de l'association positive observée, s'accroît avec l'augmentation du nombre de cigarettes consommées chaque jour. À la figure 5.2, on remarque aussi que, plus la personne était jeune lorsqu'elle a commencé à fumer, plus ce risque est élevé. Ainsi, on peut affirmer qu'un fumeur qui a commencé à fumer à 14 ans présente environ 19 fois plus de chances de mourir d'un cancer du poumon qu'un non-fumeur du même âge. De même, un fumeur qui consomme environ un paquet de cigarettes par jour a 17 fois plus de chances de mourir d'un cancer du poumon qu'un non-fumeur du même âge.

FIGURE 5.1 ■ Le risque relatif (RR) de mortalité par cancer du poumon en fonction du nombre de cigarettes consommées chaque jour

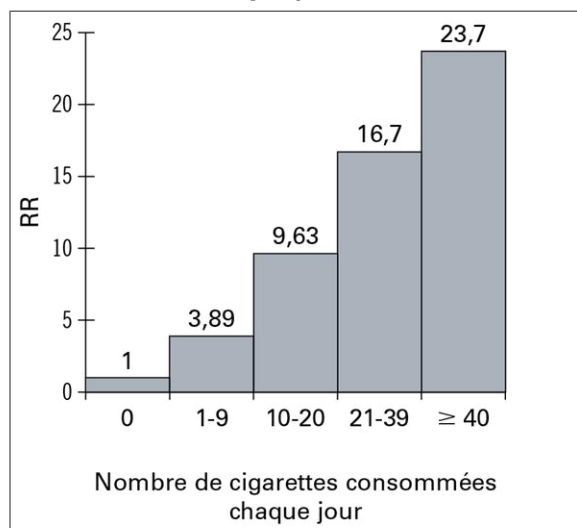
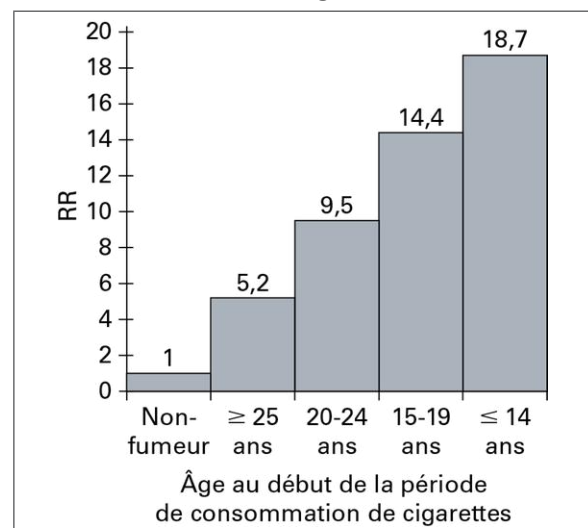


FIGURE 5.2 ■ Le risque relatif (RR) de mortalité par cancer du poumon en fonction de l'âge au début de la période de consommation de cigarettes



5.1.3 Le rapport de cotes (RC)

Le **rapport de cotes (RC)** est uniquement utilisé dans le cadre des études à visée étiologique cas-témoins. Il permet d'établir le rapport entre la cote d'exposition chez les personnes atteintes de la maladie (les cas) et la cote d'exposition chez les personnes qui n'en sont pas atteintes (les témoins).

Danielle s'est demandé si le fait de souffrir d'allergies avait une influence sur l'apparition de l'acné. Elle a donc sélectionné au hasard 40 filles qui avaient fait de l'acné pendant l'étude (les cas) et 40 autres qui n'en avaient pas fait (les témoins), pour ensuite comparer leurs prédispositions allergiques. Comme elle avait pris soin de recueillir une anamnèse complète pour toutes les adolescentes participant à l'étude, il lui a été facile de repérer, à partir des dossiers, celles qui souffraient de rhume des foins, d'asthme ou de toute autre manifestation allergique. Elle se trouvait ainsi à mener une étude cas-témoins au sein de la cohorte constituée par les jeunes filles. Le tableau 5.3 résume ses observations.

TABLEAU 5.3 ■ L'acné chez des adolescentes de 11 ans à 15 ans, suivies pendant 3 ans, en fonction de la présence de manifestations allergiques (A+) ou non (A-)

	A+	A-	Total
A fait de l'acné (cas)	25	15	40
N'a pas fait d'acné (témoins)	5	35	40
Total	30	50	80

Il importe de noter que le fait de choisir un nombre particulier de cas et de témoins a une influence directe sur le calcul du risque. En effet, le rapport entre le nombre de personnes allergiques qui ont fait de l'acné (25) et celui de l'ensemble des personnes allergiques (30) est largement fonction du nombre de cas et de témoins que Danielle aura choisi d'inclure dans l'étude. Un risque relatif calculé sur cette base ne serait pas adéquat. Par contre, les statisticiens ont établi que le rapport de cotes estimait assez correctement le risque relatif. En effet, on a vu à la section 3.2.4 que la cote est un ratio. Au numérateur de ce rapport se trouve le nombre de cas, et au dénominateur figure le nombre de témoins. Ainsi, la cote des adolescentes allergiques est ici de 25/5, et celle des adolescentes non allergiques de 15/35 :

$$\frac{25/5}{15/35}, \text{ soit } 11,67.$$

La cote peut aussi être exprimée par le rapport entre le nombre d'exposés sur le nombre de non-exposés chez les cas ou chez les témoins. Le rapport entre la cote d'exposition chez les cas 25/15 et la cote d'exposition chez les témoins 5/35 donnerait le même résultat, soit 11,67. Enfin, on aurait pu aussi calculer $(25 \times 35)/(15 \times 5)$ qui donnerait le même résultat, soit 11,67.

Le rapport de cotes, de même que le rapport de taux, s'interprète comme le risque relatif. Lorsque le rapport est différent de 1, c'est qu'il existe une association entre le facteur étudié et l'effet observé. Si le rapport est plus petit que 1, cela signifie que la

cote pour les personnes exposées au facteur est moins grande que pour les personnes non exposées. Comme pour le rapport de risques, le facteur «protège» en quelque sorte les personnes qui y ont été exposées. Par contre, si le rapport est plus grand que 1, cela indique que la cote pour les personnes exposées au facteur est plus grande que celle pour les personnes non exposées. Le facteur augmente par conséquent la probabilité d'observer l'effet chez les personnes exposées. Et plus la valeur du RC est éloignée de 1 (valeur nulle), plus forte est l'association observée.

Danielle a donc établi que le rapport de cotes, qui estime le risque de faire de l'acné si l'on souffre d'allergies, est de 11,67.

Après avoir évalué la part du hasard et celle des autres erreurs qui auraient pu fausser les mesures obtenues, Danielle en conclut qu'il existe une association positive indiquant que les personnes allergiques ont environ 12 fois plus de chances de faire de l'acné que celles qui ne le sont pas.

Les données qui servent au calcul des diverses mesures d'association examinées ci-dessus sont habituellement représentées dans ce que l'on appelle un «tableau croisé» ou «tableau de contingence», dont le tableau 5.4A constitue en quelque sorte un «modèle» et dont le tableau 5.1 illustre un exemple.

Les mesures de taux d'incidence sont aussi généralement présentées sous forme de petit tableau, dont le tableau 5.4B constitue en quelque sorte un «modèle» et dont le tableau 5.2 montre un exemple.

Les calculs des différentes mesures d'association à partir des données représentées dans un tableau de contingence s'effectuent au moyen des formules présentées sous chacun de ces tableaux.

TABLEAU 5.4 ■ Tableaux croisés (ou de contingence) représentant les données qui servent au calcul des diverses mesures d'association

A. Représentation des proportions ou des cotes				B. Représentation des taux d'incidence			
	Exposé	Non exposé	Total		Exposé	Non exposé	Total
Malade	<i>a</i>	<i>b</i>	<i>a + b</i>	Malade	<i>a</i>	<i>b</i>	<i>a + b</i>
Non malade	<i>c</i>	<i>d</i>	<i>c + d</i>	Expérience	T_1	T_0	$T_1 + T_0$
Total	<i>a + c</i>	<i>b + d</i>	<i>a + b + c + d</i>				

$$\text{Risque attribuable: } RA = \frac{a}{a+c} - \frac{b}{b+d}$$

$$\text{Risque relatif (proportions): } RR = \frac{(a/a+c)}{(b/b+d)}$$

$$\text{Rapport de cotes: } RC = \frac{a/b}{c/d} \text{ ou } \frac{a/c}{b/d}$$

$$\text{Risque attribuable: } RA = (a/T_1) - (b/T_0)$$

$$\text{Risque relatif (taux d'incidence): } RR = \frac{(a/T_1)}{(b/T_0)}$$

5.1.4 Le risque relatif à la population (RRP)

Roxanne a pu obtenir le taux de mortalité chez l'ensemble des adultes de la ville de New York pour l'année précédente. Par ailleurs, grâce à son doigté et à ses nombreux contacts, elle connaît aussi le nombre de décès survenus dans la communauté des personnes qui font usage de drogues intraveineuses (UDI) au cours de la même année. Elle s'est demandé si le taux de mortalité dans cette communauté était comparable à celui de la population adulte dans son ensemble. Sur le plan épidémiologique, ce que Roxanne veut faire, c'est comparer le risque de mourir dans la communauté des UDI à celui de mourir dans la population en général. Il s'agit donc d'une extension de la notion de risque relatif; au lieu de comparer le risque dans deux groupes, les exposés et les non-exposés, elle compare le risque dans un groupe particulier au risque connu dans la population à laquelle appartient ce groupe. En anglais, cette mesure d'association porte le nom de *Standardized Mortality (or Morbidity) Ratio (SMR)*, une appellation qui est même utilisée dans des textes publiés en d'autres langues. Dans le présent ouvrage, on désignera cette mesure par le nom de **risque relatif à la population (RRP)**.

Le RRP est utile chaque fois que l'on connaît un risque d'un événement (mort ou maladie) dans un groupe exposé à un facteur, mais en l'absence d'un groupe de comparaison (non exposé à ce facteur) qui rendrait possible le calcul d'un risque relatif simple. Si l'on connaît le risque que cet événement se produise dans la population dont est issu le groupe observé, on pourra arriver à cerner le risque relatif (RR). Le RRP sera d'autant plus proche du RR que la taille de la population sera grande comparativement à celle du groupe considéré et que le RR ne sera pas trop élevé³.

Dans la pratique, on connaît généralement le nombre de cas ou de décès dans le groupe exposé au facteur étudié et il est relativement aisé de calculer l'incidence cumulée de l'événement. Par contre, la population générale est une population dynamique ouverte, l'incidence cumulée ne s'y calcule pas facilement. On supposera donc que le taux d'incidence (I) a été constant pendant la période d'observation (t) et que le taux d'incidence multiplié par le temps d'observation est faible. Dans ces conditions, l'incidence cumulée se calcule avec la formule $IC = I \times \Delta t$.

Si l'on applique ce risque au groupe observé, on obtient le nombre de décès ou de malades attendus, c'est-à-dire le nombre de décès ou de cas qui auraient été observés si la mortalité ou la morbidité dans la population étudiée (le groupe) avait été la même que dans la population générale (celle dont est issue la population étudiée). En définitive, établir le rapport des deux risques consiste à déterminer le rapport entre les événements observés et les événements attendus⁴:

$$\text{RRP} = \frac{\text{événements observés}}{\text{événements attendus}} = \frac{o}{A}$$

EXEMPLE 5.3

Roxanne a recensé 217 décès dans la communauté des UDI sur une période de un an. Cette communauté comptait environ 10 000 personnes en milieu d'année. Pour la même année, à New York, la mortalité (M) chez les adultes était de $10,2 \times 10^{-3} \text{ an}^{-1}$. En postulant que le taux de mortalité a été stable cette année-là et que $M \times \Delta t$ était faible, le risque de mourir calculé sur un an serait de $10,2 \times 10^{-3}$. Le nombre de décès attendus (A) s'exprimerait donc comme suit :

$$A = 10\,000 \text{ personnes} \times \frac{10,2}{1000}, \text{ soit } 102 \text{ décès.}$$

Quant au risque relatif à la population, il serait de :

$$\text{RRP} = \frac{217 \text{ décès}}{102 \text{ décès}}, \text{ soit } 2,13.$$

Roxanne en conclut que le risque de décès était à peu près deux fois plus élevé dans la communauté des UDI que dans la population adulte de la ville de New York cette année-là.

Les mesures d'association et les types d'études

On a vu que, contrairement aux études à visée étiologique transversales ou longitudinales de cohorte(s), les études à visée étiologique cas-témoins ne permettent pas de calculer des mesures de fréquence de la maladie ou des décès chez les exposés et les non-exposés parce que le nombre de cas et de témoins déterminés par le chercheur influencera directement ce calcul. Ainsi, les différentes mesures d'association doivent être utilisées dans des contextes d'étude précis. Le tableau 5.5 résume les mesures d'association pouvant être employées selon les différents types d'études à visée étiologique.

TABLEAU 5.5 ■ Les mesures d'association selon le type d'étude à visée étiologique

Type d'étude à visée étiologique	Type de mesures d'association
Transversale et longitudinale ou de cohorte(s)	<ul style="list-style-type: none"> • Risque attribuable (RA) • Risque relatif (RR) : rapport de proportions ; rapport de taux
Cas-témoins	<ul style="list-style-type: none"> • Rapport de cotes (RC)

5.1.5 Le coefficient de corrélation linéaire

Les risques attribuable et relatif et les rapports de cotes mesurent l'association entre des variables catégorielles. Pour mesurer l'association entre des variables quantitatives continues, il faut recourir à une autre technique : le **coefficient de corrélation linéaire**.

EXEMPLE 5.4

Mounkaïla, un collègue épidémiologiste de l'omnipraticienne Lisa, a effectué une étude pour la compagnie d'assurances Viéternelle. L'un des propriétaires de cette compagnie est revenu d'un voyage au Swaazi, il y a quelques années, avec la conviction que la racine de «youmiyoum» abaissait la pression artérielle de façon notable. Comme il a été démontré qu'il existe une relation entre la pression artérielle diastolique élevée et le décès précoce, il a réussi à convaincre ses associés de mener une vigoureuse campagne de promotion auprès des clients de la compagnie. Celle-ci fournissait même le «youmiyoum» gratuitement à ceux qui en faisaient la demande, en leur indiquant que la dose recommandée était de 200 mg/jour. Cependant, il était bien connu que le goût amer de la racine en rebutait plusieurs.

Après trois ans, le conseil d'administration a voulu savoir si ses efforts étaient rentables. À titre de fournisseur, il connaissait la quantité fournie et prétendument consommée par chacun de ses clients. Il connaissait aussi la pression artérielle diastolique moyenne de ceux-ci avant le début de la campagne de promotion et après ces trois années ; la pression diastolique moyenne avait en fait baissé d'environ 3 mm Hg pendant cette période. Le conseil d'administration a donc demandé à Mounkaïla s'il existait une association entre la quantité de «youmiyoum» consommée pendant les trois années en question et la différence de pression artérielle diastolique observée. Mounkaïla présente les résultats de son étude aux membres du conseil d'administration.

«L'étude a montré que, au cours des trois dernières années, la pression artérielle diastolique des clients a diminué de 2,8 mm Hg en moyenne avec un écart type de 1,2 mm Hg. Pendant cette même période, la consommation moyenne de «youmiyoum» a été de 64,7 g par client avec un écart type de 22,4 g.»

La signification du coefficient de corrélation

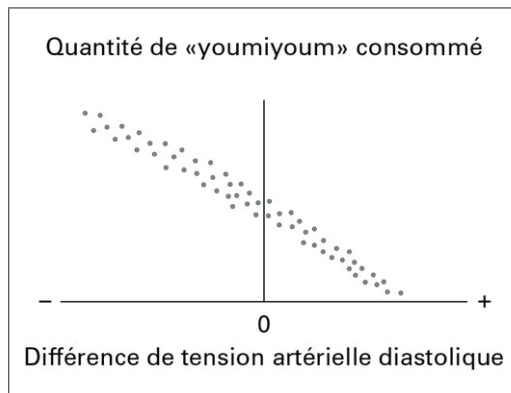
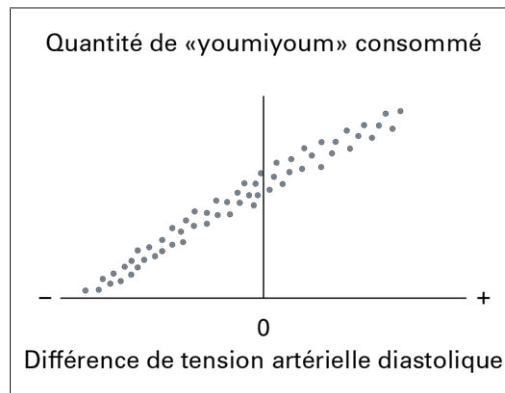
Pour vérifier l'association entre la consommation de «youmiyoum» et la modification de la pression artérielle, j'ai calculé le coefficient de corrélation linéaire entre ces deux variables. Il faut savoir que le coefficient de corrélation linéaire « r » nous renseigne sur l'intensité de la relation linéaire entre deux variables quantitatives et sur le sens de cette relation. Si la relation est forte, le coefficient de corrélation, en valeur absolue, sera près de 1, ce qui démontrera une relation entre la consommation de «youmiyoum» et la modification de la pression artérielle. Si, en revanche, la relation est faible, le coefficient, toujours en valeur absolue, se situera près de 0. Dans ce cas, il n'y aura pas de relation entre la consommation de «youmiyoum» et la modification de la pression artérielle. Par ailleurs, une relation directement proportionnelle donnera un coefficient de corrélation positif. Cela traduira une hausse de la pression artérielle correspondant à l'augmentation de la consommation de «youmiyoum». Une relation inversement proportionnelle donnera un coefficient négatif, qui exprimera plutôt le résultat recherché, soit une diminution de la pression artérielle associée à l'augmentation de la consommation de «youmiyoum». Le tableau 5.6 résume les valeurs possibles du coefficient de corrélation linéaire.

TABLEAU 5.6 ■ Les valeurs du coefficient de corrélation linéaire « r »

Relation	Forte	Faible
Proportionnelle	Coefficient r près de 1	Coefficient r près de 0
Inversement proportionnelle	Coefficient r près de -1	Coefficient r près de 0

Un coefficient négatif Voyons comment cela s'applique dans le cas qui nous occupe. L'hypothèse de départ de votre compagnie était que plus les personnes ingèreraient de «youmiyoum», plus leur pression artérielle diastolique diminuerait. C'est donc dire que la modification de la pression artérielle devait être inversement proportionnelle à la consommation de «youmiyoum». Si cette hypothèse avait été vérifiée, nous aurions observé un coefficient de corrélation linéaire près de -1 entre la consommation de «youmiyoum» et la différence de pression artérielle observée (pression après un an de consommation - pression au début). La représentation graphique de ce phénomène aurait donné la figure 5.3.

Un coefficient positif «Si, par contre, à l'inverse de vos attentes, il y avait eu une relation linéaire directe entre le «youmiyoum» et la pression artérielle diastolique, nous aurions observé un coefficient de corrélation linéaire près de 1, ce qui aurait signifié que plus on consommait de «youmiyoum», plus la pression artérielle diastolique s'élevait. Graphiquement, on aurait obtenu la figure 5.4.

FIGURE 5.3 ■ La relation linéaire inversement proportionnelle correspondant à un coefficient de corrélation près de -1**FIGURE 5.4 ■ La relation linéaire proportionnelle correspondant à un coefficient de corrélation près de 1**

Un coefficient près de 0 «Je vous fais grâce des calculs effectués au cours de l'étude. Je me contenterai de vous dire qu'ils m'ont donné un coefficient de corrélation linéaire de 0,07; un résultat bien près de 0. J'en déduis qu'il n'existe pas de corrélation linéaire entre la consommation de «youmiyoum» et la pression artérielle. Devant un tel résultat, deux interprétations sont possibles. Dans le premier cas, il n'y a aucune

corrélation entre le «youmiyoum» et la pression artérielle diastolique, ce qui revient à dire que la consommation de cette plante n'a aucun effet sur la pression artérielle. Dans le second cas, il y a une corrélation, mais elle n'est pas linéaire.»

Examinons cette seconde possibilité. En effet, on peut penser que, jusqu'à l'atteinte d'une certaine dose, la pression artérielle diminue et que, une fois cette dose «dépassée», elle augmente de nouveau. On peut comprendre intuitivement que, dans un tel cas, les deux tendances opposées s'annulent, ce qui explique qu'on obtienne un coefficient de corrélation bien près de 0. La figure 5.5 illustre cette situation.

Malheureusement, le graphique illustrant la relation entre la quantité de «youmiyoum» consommé et la différence de pression artérielle diastolique établie en fonction des résultats de notre étude se présente plutôt comme l'illustre la figure 5.6.

FIGURE 5.5 ■ La relation non linéaire correspondant à un coefficient de corrélation linéaire près de 0

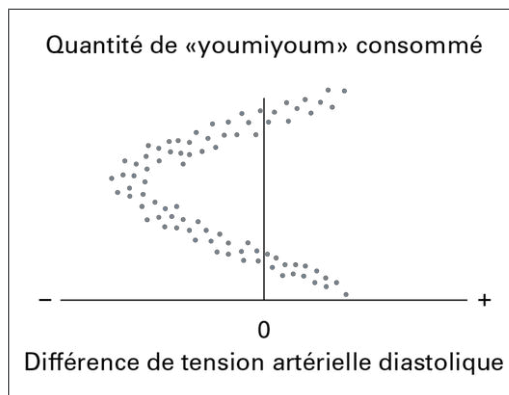
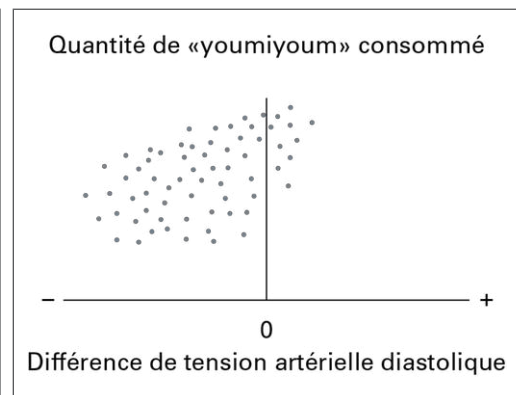


FIGURE 5.6 ■ L'absence de relation entre les variables correspondant à un coefficient de relation près de 0



« Comme vous pouvez le constater, il n'y a aucune relation entre les deux variables étudiées. La baisse de la pression artérielle diastolique moyenne est donc due à une cause autre que la consommation de «youmiyoum». »

À la lumière de ce résultat, la campagne de promotion fut immédiatement interrompue, et le «youmiyoum» tomba dans l'oubli le plus absolu. Ce qui explique d'ailleurs pourquoi c'est la première et la dernière fois que vous en entendez parler!

5.2 La causalité

Au cours d'une réunion de l'équipe de la polyclinique, Danielle présente les premiers résultats de l'étude sur le chiogène hispide. À l'aide de techniques qui sont abordées dans le chapitre 11, elle a déterminé que l'association entre la plante et la diminution

de l'acné n'est pas due au hasard, mais qu'elle se révèle bel et bien significative sur le plan statistique. Elle émet donc l'opinion que le chiogène hispide cause effectivement une diminution de l'acné. Un dialogue s'engage entre Lisa, sceptique, et Danielle.

5.2.1 L'association, la temporalité et la conséquence

Lisa : « Danielle, qu'est-ce qui t'amène à parler de relation causale dans ce cas-ci ? L'association est certaine sur le plan quantitatif, mais ce n'est pas parce qu'on a une association qu'il s'agit nécessairement d'une cause. Par exemple, il y a probablement une association très forte entre la présence de la lune dans le ciel et la consommation d'électricité pour éclairer les routes : les deux surviennent bien plus souvent la nuit que le jour. Il n'y a pourtant aucune relation de cause à effet entre les deux phénomènes. On n'éclaire pas moins les rues quand la lune est pleine, et les pannes d'électricité ne la font pas disparaître. »

Danielle : « Comme tu le soulignes à juste titre, Lisa, la découverte d'une association à l'aide d'une méthode d'observation quantitative ne permet pas d'établir un lien de cause à effet entre deux phénomènes. D'autres critères d'analyse doivent être vérifiés. La temporalité, ou chronologie, constitue un autre critère important. La cause doit toujours précéder l'effet. Dans notre étude, les épisodes d'acné diminuent après la consommation de "petit thé". Cependant, les critères de temporalité et d'association ne sont pas à eux seuls suffisants pour parler de relation causale.

« Il faut un critère supplémentaire pour parler de **causalité**. Ce critère, je l'appelle la "conséquence". Il faut pouvoir démontrer qu'un changement à la cause implique un changement de l'effet.

« Il importe de noter que cette notion de "conséquence" implique obligatoirement une relation asymétrique, et non réciproque. Une relation asymétrique existe, par exemple, quand je plonge des macaronis secs et durs dans l'eau bouillante, et que cette eau entraîne l'amollissement des macaronis. Quoi que je fasse après la cuisson, je ne peux redonner à mes macaronis leur texture originale. En revanche, une relation réciproque existe dans le cas d'une personne dépressive qui présente également des symptômes d'anxiété. L'anxiété peut être entraînée par la dépression ; en même temps, la dépression peut être exacerbée par l'anxiété. »

5.2.2 Les critères de la causalité

Pour tenter de cerner la conséquence d'un facteur sur un autre, plusieurs auteurs ont en fait proposé une série de postulats, ou critères. Entre autres, Robert Koch, qui avait démontré le rôle du bacille tuberculeux dans la tuberculose, proposa en 1882 quatre postulats permettant de reconnaître un agent pathogène⁵.

- Le « corps étranger » (ou microorganisme) doit toujours être trouvé au moment où la maladie est présente.
- On doit pouvoir montrer, en l'isolant et en le cultivant, que le « corps étranger » est un organisme vivant distinct des organismes qu'on trouve par ailleurs chez le malade.

- Le « corps étranger » qu'on trouve chez le malade doit être en relation avec les lésions que provoque la maladie.
- Le « corps étranger » cultivé sur plusieurs générations doit provoquer la maladie lorsqu'il est inoculé à des animaux chez qui elle est susceptible de se développer.

Ces critères sont cependant trop restrictifs pour être utilisés lorsqu'on recherche la cause des maladies non infectieuses. C'est ce qui a motivé Bradford Hill⁶, en 1965, dans le contexte des débats concernant la part attribuable à la cigarette dans le cancer du poumon, à proposer les neuf critères de causalité suivants.

1. Temporalité. L'exposition au facteur de risque doit précéder l'apparition de la maladie.
2. Force de l'association. Elle se mesure à l'aide du risque relatif, du rapport de cotes et des analyses de corrélation. Plus la valeur de la mesure est élevée, plus forte est l'association.
3. Constance et validité des résultats. Une association est constante si plusieurs études menées auprès de populations différentes la confirment. Ces études doivent présenter une bonne validité interne et externe.
4. Spécificité. Ce critère suggère une relation spécifique entre le facteur de risque et la maladie. Si un facteur de risque est constamment et uniquement observé dans le contexte de l'apparition d'une maladie, il serait plausible de croire à une relation de cause à effet.
5. Relation dose-réponse. L'augmentation de la durée ou de l'intensité de l'exposition au facteur de risque devrait entraîner une augmentation de la fréquence ou de la gravité de la maladie; c'est-ce qu'on appelle le « gradient biologique ».
6. Vraisemblance. L'association observée devrait présenter une bonne concordance avec les connaissances générales du phénomène en cause.
7. Cohérence avec les connaissances antérieures. L'association observée et son effet devraient être corroborés par les études antérieures. Idéalement, les études expérimentales hasardisées fournissent les meilleures preuves.
8. Plausibilité biologique. Le phénomène observé devrait trouver une explication associée aux connaissances de la biologie humaine.
9. Analogie. Des similitudes avec d'autres expositions et leurs effets connus devraient être établies.

On notera que le critère de **relation dose-réponse** se trouve respecté lorsque l'importance de la maladie est en rapport avec l'intensité du facteur qui la cause. Ainsi, il a été démontré que le risque de cancer du poumon est d'autant plus élevé que l'exposition à la fumée de cigarette s'avère importante, soit au chapitre de la durée (plus on a fumé longtemps, plus risque qu'un cancer se développe), soit à celui de la quantité (plus le nombre de cigarettes fumées dans une journée est élevé, plus le risque qu'un cancer se développe le devient aussi).

Danielle: « Ces critères de causalité des maladies ont par la suite été rediscutés et maintenus avec des nuances par un grand nombre d'auteurs en épidémiologie⁷.

Par ailleurs, ces critères ne sont pas absolus; les chercheurs peuvent les adapter aux différents contextes et objets d'étude. En effet, dans le domaine de la causalité, la recherche se poursuit et continuera de se poursuivre, puisque tout progrès scientifique est continuellement remis en question par l'acquisition de connaissances nouvelles.

«Examinons maintenant les résultats de l'étude que nous avons menée sur le “petit thé” en nous servant des critères applicables de Hill, afin de vérifier si l'association révélée par l'approche quantitative permet de conclure à une relation de cause à effet.

La temporalité

«Vous connaissez le protocole “à double insu” que nous avons utilisé et nous nous accordons, je crois, pour dire que cette étude est valide. En outre, le critère de temporalité est respecté parce que le protocole d'étude nous permet de démontrer que la cause précède toujours l'effet.

La force de l'association

«Nous avons déjà vu que l'association entre la consommation de chiogène hispide et la diminution de l'acné est significative et assez forte.

La vraisemblance et la plausibilité biologique

«Il n'est pas invraisemblable que le chiogène hispide ait joué un rôle dans l'effet observé. Il est possible d'envisager que l'une de ses composantes, que nous ne connaissons pas encore, puisse renforcer l'épiderme et le rendre moins sensible aux lésions acnéiformes. Cela n'entre pas en contradiction avec nos connaissances antérieures.

L'analogie

«Par analogie, on peut penser aux personnes chez qui l'apport en niacine (une vitamine du complexe B) est déficient. Ces personnes présentent une grave affection cutanée qui guérit rapidement lorsque le déficit est corrigé. Les adolescents qui font de l'acné souffrent peut-être d'un déficit relatif quelconque que le chiogène hispide viendrait compenser.

La spécificité

«La spécificité, par contre, reste à démontrer. En effet, le chiogène hispide est peut-être efficace dans le traitement d'autres affections et l'acné est peut-être tout aussi efficace, voire plus efficace, traité et prévenu avec un autre agent thérapeutique.

La relation dose-réponse

«Enfin, la relation dose-réponse n'a pas été confirmée et l'étude ne peut à proprement parler prétendre à la plausibilité biologique, puisque l'ingrédient actif du chiogène hispide demeure inconnu.»

Lisa et Danielle concluent que les résultats de l'étude sont intéressants, notamment parce que le chiogène hispide diminue le risque d'acné de 15 % (1 - 0,85). Il agit donc sur une partie des déterminants de l'acné.

Elles ajoutent: «Il est certain que des recherches supplémentaires sont nécessaires. Cette première étude suscite certes de l'intérêt compte tenu de la prévalence importante de l'acné chez les adolescents. Il faudrait mettre dans le coup des biologistes, des botanistes et des pharmaciens pour déceler l'ingrédient actif du chiogène hispide. Il faudrait aussi effectuer une étude plus approfondie pour confirmer la relation dose-réponse et éventuellement permettre d'ajouter le chiogène hispide à notre arsenal thérapeutique...»

5.3 La mesure de l'impact d'un facteur d'exposition sur la fréquence des maladies et des décès

Le risque relatif et le rapport de cotes permettent de vérifier la présence d'une association et d'en déterminer la force. Lorsque cette association est vérifiée, l'étape suivante consiste à déterminer quelle est la contribution de l'exposition étudiée sur l'augmentation ou la diminution de la fréquence de la maladie ou des décès⁸.

5.3.1 La fraction étiologique chez les exposés

On a vu que l'impact d'un facteur d'exposition sur le nombre de nouveaux cas d'une maladie ou le nombre de décès liés à une exposition peut être mesuré grâce au risque attribuable, RA. On peut également exprimer cet impact grâce au rapport entre le risque attribuable et le risque chez les exposés. On obtient alors la proportion des nouveaux cas de la maladie ou des décès qui, chez les exposés, sont spécifiquement attribuables à l'exposition.

En reprenant les résultats de l'étude britannique sur les fumeurs citée dans l'exemple 5.1, on obtient ainsi:

$$FE_1 = \frac{0,9 - 0,07}{0,9} = \frac{83 \text{ décès par cancer FE du poumon par } 100\,000 \text{ fumeurs - année}}{0,9} \\ = 0,92, \text{ soit } 92\%$$

La **fraction étiologique chez les exposés** (FE_1) s'obtient également à l'aide du RR de la façon suivante:

$$RR = \frac{0,9 \times 10^{-3} \text{ an}^{-1}}{0,07 \times 10^{-3} \text{ an}^{-1}} = 12,9 \\ FE_1 = \frac{RR - 1}{RR}, \text{ soit } \frac{12,9 - 1}{12,9} = 0,92, \text{ soit } 92\%$$

Cela permet d'estimer que, dans l'échantillon des fumeurs étudiés, 92% des nouveaux cas de cancer du poumon observés seraient attribuables au tabagisme. En d'autres mots, on peut estimer que le tabagisme est responsable de 92% des cas de cancer du poumon chez les fumeurs étudiés.

5.3.2 La fraction prévenue chez les exposés

Dans le contexte d'un effet protecteur comme celui observé dans l'exemple portant sur le « petit thé », le rapport entre le RA (10 %) et le risque chez les sujets n'ayant pas consommé de thé (65 %), soit $10\%/65\% = 15,4\%$, permet d'estimer que la consommation de chiogène hispide permettrait d'éviter 15,4 % des cas d'acné qui surviennent chez les sujets qui n'en consomment pas. En d'autres mots, 15,4 % des cas potentiels d'acné ont pu être évités par la prise de chiogène hispide chez les consommateurs de « petit thé ». C'est ce qu'on appelle la **fraction prévenue chez les exposés**.

5.3.3 La fraction étiologique et la fraction prévenue totales

Il est également possible de calculer la fraction étiologique dans la population totale. Cette dernière étant composée d'individus exposés au facteur étudié et d'individus non exposés à ce facteur, nous devons alors « pondérer » le calcul de la fraction étiologique en tenant compte du poids relatif des individus exposés et des individus non exposés. En fait, dans une population, le risque de survenue d'une maladie variera inévitablement en fonction de la proportion d'individus exposés au facteur de risque susceptible de provoquer cette maladie. Ainsi, le calcul de la **fraction étiologique totale** passe d'abord par celui du risque total.

EXEMPLE 5.5

En supposant que, dans l'étude sur la relation entre le tabagisme et le cancer du poumon, 40 % des médecins étaient des fumeurs et 60 % des non-fumeurs, le risque exprimé par le taux de mortalité par cancer du poumon étant de 7 décès pour 100 000 personnes-année chez les non-fumeurs et de 90 décès pour 100 000 personnes-année chez les fumeurs, la fraction étiologique totale s'obtiendrait de la façon suivante.

Étape 1: Calcul du risque total

$$R_t = (0,4 \times 90) + (0,6 \times 7) = 40,2, \text{ soit } 40,2 \text{ décès pour } 100\,000 \text{ personnes-année}$$

Risque chez les exposés Risque chez les non-exposés

Proportion d'individus exposés Proportion d'individus non exposés

Étape 2: Calcul de la fraction étiologique totale

Pour les études à visée étiologique de cohorte(s) comme dans le cas présent :

$$FE = \frac{R_t - R_0}{R_t}, \text{ soit } \frac{40,2 - 7}{40,2} = 0,826, \text{ soit } 82,6\%$$

Ce résultat permet d'estimer que, dans la population totale d'où est extrait l'échantillon des médecins, 82,6 % des décès par cancer du poumon seraient attribuables au tabagisme. Ce résultat permet également d'estimer que 82,6 % des décès par cancer du poumon auraient pu être évités si le tabagisme n'avait pas été présent dans cette population.

Pour les études à visée étiologique cas-témoins, la fraction étiologique totale s'obtient de la façon suivante :

$$FE_t = pc_1 \times \frac{RC - 1}{RC}$$

où pc_1 représente la proportion des cas exposés parmi l'ensemble des cas.

Dans le contexte d'une association protectrice où le risque chez les exposés est plus petit que le risque chez les non-exposés, on calculera selon la même approche une **fraction prévenue totale**.

5.3.4 Le nombre de personnes à traiter (NPT)

Dans une étude expérimentale où l'on compare l'effet d'un traitement à un placebo, il sera utile de mesurer le nombre de patients qu'il faudrait traiter pour prévenir un nouveau cas de la maladie. Le nombre de patients à traiter pour prévenir un nouveau cas s'obtient en calculant l'inverse du risque attribuable : $-1 / RA$.

EXEMPLE 5.6

Nous avons vu dans le tableau 5.1, avec l'étude expérimentale sur la consommation du «petit thé», que le risque d'être atteint d'au moins un épisode d'acné est de 55 % dans le groupe ayant consommé le «petit thé» alors qu'il est de 65 % dans le groupe ayant consommé un placebo. Le RA est de 55 % – 65 %, soit -10 %. Autrement dit, la réduction du risque grâce au «petit thé» est de 10 % (65 % – 55 %). Le nombre requis de personnes à traiter (NPT) correspond à $-1/-10\%$, ce qui donne 10. Il faudrait donc traiter 10 patients pour prévenir un nouveau cas d'épisode d'acné. On doit noter ici que la durée de traitement pourrait faire varier ce résultat. Le NPT est donc interprétable uniquement si l'on connaît la durée du traitement qui est ici de trois ans.

En résumé

Le risque attribuable, le risque relatif, le rapport de cotes, le risque relatif à la population et le coefficient de corrélation linéaire sont des mesures d'association. Quand la mesure de base est une prévalence, une incidence cumulée ou un taux d'incidence, la mesure d'association correspondante est le risque relatif. On se situe alors dans le contexte d'études à visée étiologique de cohorte(s). Dans une étude cas-témoins, on estimera le risque relatif à partir du rapport de cotes.

Le risque relatif à la population établit le rapport entre le risque des événements observés

dans le groupe étudié et ce même risque estimé dans la population générale. Le coefficient de corrélation linéaire permet de mesurer l'association linéaire entre deux variables continues.

Les critères de causalité permettent quant à eux de déceler les relations causales, parce que la découverte d'une association au moyen des mesures quantitatives mentionnées précédemment permet d'établir une association positive ou négative, mais ne permet pas à elle seule d'établir une association causale.

Les mesures d'association permettent de vérifier la présence d'une association et d'en mesurer la force. Pour leur part, les mesures d'impact telles que la fraction étiologique chez les exposés, la fraction prévenue chez les exposés, la fraction étiologique dans la population totale et la fraction

prévenue dans la population totale permettent d'exprimer en pourcentage l'influence d'un facteur d'exposition sur la fréquence des maladies et des décès. Enfin, il est possible de calculer le nombre de personnes qui doivent recevoir un traitement pour obtenir un résultat.

Exercices



Voir la section complète des exercices sur le Web.

1. Pendant 20 ans, on a suivi 150 enfants ayant été exposés à de fortes concentrations de mercure pendant les 6 premiers mois de leur vie. Un autre

groupe de 200 enfants, non exposés au mercure, a aussi été suivi pendant la même période. On a observé les résultats suivants.

Troubles	Exposition au mercure +	Exposition au mercure -	Total
Troubles neurologiques +	40	8	48
Troubles neurologiques -	110	192	302
Total	150	200	350

- A)** De quel type d'étude s'agit-il?
- Une étude expérimentale.
 - Une étude à visée étiologique cas-témoins.
 - Une étude à visée étiologique de cohorte non expérimentale.
- B)** Déterminez la bonne façon de calculer le risque relatif de présenter des troubles neurologiques chez les enfants qui ont été exposés au mercure par rapport aux enfants qui n'y ont pas été exposés.
- $(40/48)/(110/302)$, soit 2,3.
 - $(40/110)/(8/192)$, soit 8,7.
 - $(40/150)/(8/200)$, soit 6,67.
2. Laquelle des mesures suivantes permet de vérifier la présence d'une association dans une étude cas-témoins et d'en quantifier la force?
- Le risque attribuable.
 - La fraction étiologique chez les exposés.
 - Le rapport de cotes.
 - Le risque relatif.
3. Laquelle des mesures suivantes permet de mesurer l'impact du facteur d'exposition étudié sur la fréquence des nouveaux cas de la maladie observés?
- Le risque attribuable.
 - La fraction étiologique.
 - Le rapport de cotes.
 - Le risque relatif.
4. Parmi les critères de causalité, la temporalité implique :
- que la durée d'exposition doit être suffisamment longue pour permettre l'apparition de la maladie.
 - que l'exposition au facteur de risque doit précéder la maladie.
 - que la durée du suivi des personnes observées doit permettre de bien mesurer les effets du facteur d'exposition étudié.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 6

L'utilité clinique des notions de précision et de validité

Collaboration à la 3^e édition :

Danielle Laurin

André Simpson

Au terme de ce chapitre, vous pourrez :

- utiliser les notions de justesse, de validité et de précision des mesures en épidémiologie ;
- juger de la validité et de l'utilité d'une publication scientifique en vérifiant la présence de biais pouvant induire des erreurs dans les résultats d'une étude.

Objectifs spécifiques	Contenu
6.1 Définir les notions de base permettant d'obtenir des mesures justes de la santé des individus et des populations.	6.1 La justesse d'une mesure 6.1.1 La validité 6.1.2 La précision 6.1.3 Les relations entre la précision et la validité
6.2 Expliquer les types de biais et leur(s) source(s), et déterminer les moyens à prendre pour prévenir et contrôler les biais potentiels.	6.2 Les types de biais 6.2.1 Le biais de sélection 6.2.2 Le biais d'information 6.2.3 Le biais de confusion





Mise en situation

Michel, 28 ans, se présente au service de consultation sans rendez-vous de la polyclinique Milo. Il se plaint de douleurs lorsqu'il urine et dit avoir observé un écoulement urétral. Ces symptômes laissent soupçonner la présence d'une infection transmissible sexuellement et par le sang (ITSS). Michel fume 15 cigarettes par jour depuis l'âge de 16 ans, mais il est par ailleurs en très bonne santé.

Lisa, l'omnipraticienne, poursuit l'anamnèse et l'examen clinique, et elle termine la démarche diagnostique en faisant les prélèvements pertinents. Elle dispense les conseils appropriés sur le tabagisme et les ITSS, puis prend la mesure de la pression artérielle de Michel. Elle obtient des résultats de 140/84 mm Hg et de 136/80 mm Hg

au moment des deux prises réalisées selon les méthodes recommandées¹.

Pertinence clinique

Que ce soit à l'anamnèse, au cours de l'examen clinique ou des examens paracliniques, la démarche clinique et épidémiologique est fondée sur la collecte d'informations de nature qualitative et quantitative. Ces mesures orientent les décisions concernant les interventions préventives, la confirmation du diagnostic et le choix du traitement. Il est donc important de connaître les notions de base sur la façon appropriée d'obtenir ces mesures et de pouvoir juger de leur qualité. Le jugement clinique et épidémiologique se fonde sur la précision et la validité des mesures.

6.1 La justesse d'une mesure

Les mesures obtenues dans un contexte clinique ou à l'occasion d'études épidémiologiques peuvent être erronées. On reconnaît généralement deux types d'erreurs: l'erreur systématique et l'erreur aléatoire. La validité a trait à l'absence d'erreurs systématiques, alors que la précision renvoie à l'absence d'erreurs aléatoires. Certains auteurs utiliseront les termes «fiabilité», «répétabilité», «fidélité» ou «reproductibilité»² pour désigner le degré selon lequel le résultat d'une mesure peut être reproduit et «précision» d'une mesure en référence à l'absence de variation attribuable à l'effet du hasard. Bernard et Lapointe³ proposent la notion de «justesse» d'une mesure: une mesure juste est à la fois valide et précise.

6.1.1 La validité

La notion de **validité** fait appel à la capacité d'appréhender correctement un phénomène ou d'en mesurer la valeur réelle. Ainsi, une mesure de la pression artérielle

est valide si elle correspond à la valeur réelle de la pression artérielle de l'individu au moment où elle est mesurée. De la même manière, une culture effectuée pendant l'examen d'une personne souffrant d'une urétrite à gonocoque (une ITSS) qui révèle la présence de ce microorganisme donne une information valide. En revanche, cette mesure n'est pas valide si la culture se révèle négative alors que la personne est effectivement infectée.

Il existe deux types de validité : interne et externe. La capacité d'une étude d'estimer correctement les mesures ou les liens recherchés relève de la **validité interne** ; la capacité de généraliser un résultat à une population cible a trait à la **validité externe**.

Michel demande à Lisa de le conseiller quant aux moyens de cesser de fumer. Dans le but de mettre à jour ses connaissances sur la question, Lisa parcourt une étude effectuée auprès d'adultes fréquentant l'université. Les auteurs de cette étude soutiennent que les résultats obtenus établissent l'efficacité d'un nouveau programme de désaccoutumance au tabac. L'étude semble bien conçue : deux groupes comparables entre eux ont été constitués. L'un de ces groupes prend part au programme en question, alors que l'autre n'y participe pas.

Cette étude expérimentale ne présente pas de problème de validité interne ; le programme s'avère réellement efficace, puisque l'on démontre un renoncement au tabac plus fréquent chez les personnes qui y ont participé.

Toutefois, compte tenu des conditions de réalisation de l'étude, les résultats ne peuvent être généralisés à l'ensemble des universitaires ni à d'autres personnes ne fréquentant pas l'université. En effet, tous les étudiants sélectionnés pour participer à l'étude étaient des hommes âgés de 35 à 40 ans qui fumaient tous moins de 10 cigarettes par jour au début de l'étude. Celle-ci présente donc des problèmes de validité externe. Dans ce contexte, Lisa ne sera probablement pas en mesure de conseiller à Michel de suivre ce programme sur la base de cette étude.

Pour juger de la valeur d'une information, il faut connaître les facteurs pouvant conduire à des erreurs dans l'estimation d'une mesure. Lorsque Lisa prend la pression artérielle de Michel, elle se demande si cette mesure est valide. Quelles sont les sources d'erreurs systématiques, ou **biais**, qui peuvent invalider cette mesure ?

Un biais apparaît lorsqu'une mesure s'écarte ou dévie systématiquement de la valeur réelle. Un appareil mal calibré, l'utilisation d'un critère d'observation erroné, le manque de collaboration du sujet si un examen exige sa participation ou une mauvaise méthode de lecture d'un appareil mènent tous à une erreur systématique. On dira d'une mesure obtenue dans ces conditions qu'elle comporte un biais. Celui-ci nuit à la validité d'une mesure ; une mesure biaisée n'est pas valide, car elle ne correspond pas à la valeur réelle. Un appareil mal calibré qui donne systématiquement des mesures de pression artérielle supérieures de 20 mm Hg à la valeur réelle ne permet donc pas d'obtenir une mesure valide. Le chercheur peut prévenir les biais en se donnant un protocole d'étude bien conçu, et le clinicien, en optant pour une démarche rigoureuse. Par ailleurs, un appareil bien calibré peut donner des mesures valides, mais peu précises, c'est-à-dire souffrant de variations dites « aléatoires ».

6.1.2 La précision

On dit d'une mesure qu'elle est précise lorsqu'elle fluctue peu. En effet, la **précision** se caractérise par l'absence relative d'erreur aléatoire.

EXEMPLE 6.1

Karine, médecin en formation, fait un stage à la polyclinique Milo. Elle a bien voulu se prêter à l'exercice suivant: une infirmière expérimentée a pris sa pression artérielle à 20 reprises: 10 fois à l'aide du sphygmomanomètre A, et 10 fois à l'aide du sphygmomanomètre B. Le tableau 6.1 présente les résultats de ces mesures obtenues avec les deux appareils.

TABLEAU 6.1 ■ La pression de Karine à l'aide des sphygmomanomètres A et B

Sphygmomanomètre A		Sphygmomanomètre B	
140,1	140,3	144,5	138,3
140,0	139,6	137,5	142,3
140,4	139,8	138,6	140,3
139,9	140,3	139,4	139,6
139,7	140,1	142,7	136,8

A – Moyenne et écart type: $140,0 \pm 0,3$ mm Hg B – Moyenne et écart type: $140,0 \pm 2,5$ mm Hg

Pour les besoins de l'exercice, Lisa considère que la pression artérielle systolique de Diane n'a pas varié au moment des mesures. Elle remarque que celles-ci fluctuent autour de la moyenne de 140 mm Hg, mais que les écarts sont beaucoup moins marqués avec l'appareil A. Si la pression artérielle de Diane était mesurée un plus grand nombre de fois, on pourrait observer une tendance des valeurs à se répartir également autour de la valeur réelle; environ 50 % de ces mesures seraient inférieures à la moyenne, et environ 50 % seraient supérieures. Lisa peut donc affirmer que cette fluctuation est aléatoire ou due au hasard. Puisqu'une mesure précise est caractérisée par l'absence relative d'erreur aléatoire, les variations plus faibles autour de la moyenne obtenues avec le sphygmomanomètre A (écart type: 0,3 mm Hg) indiquent que cet appareil est plus précis que le sphygmomanomètre B (écart type: 2,5 mm Hg).

Afin d'obtenir une mesure précise, il faut connaître les sources de variations de cette mesure. En premier lieu, les variations biologiques propres à chaque individu, qui se caractérisent par des fluctuations en fonction du moment de la journée, de l'état d'éveil, de l'absorption de nourriture ou encore des changements d'humeur, pourront être à l'origine des variations de la mesure. En outre, lorsque des mesures sont effectuées dans un groupe d'individus, elles pourront varier en raison des facteurs biologiques propres à chacun d'eux, mais aussi à cause de leurs habitudes de vie différentes (nutrition, sommeil, activités physiques), des milieux de vie distincts

(environnement, milieu de travail, etc.) et de facteurs génétiques ou pathologiques. Enfin, certaines variations pourront être attribuables aux instruments de mesure utilisés (précision intrinsèque de l'appareil), à l'observateur ou au sujet, si le protocole d'examen requiert sa collaboration.

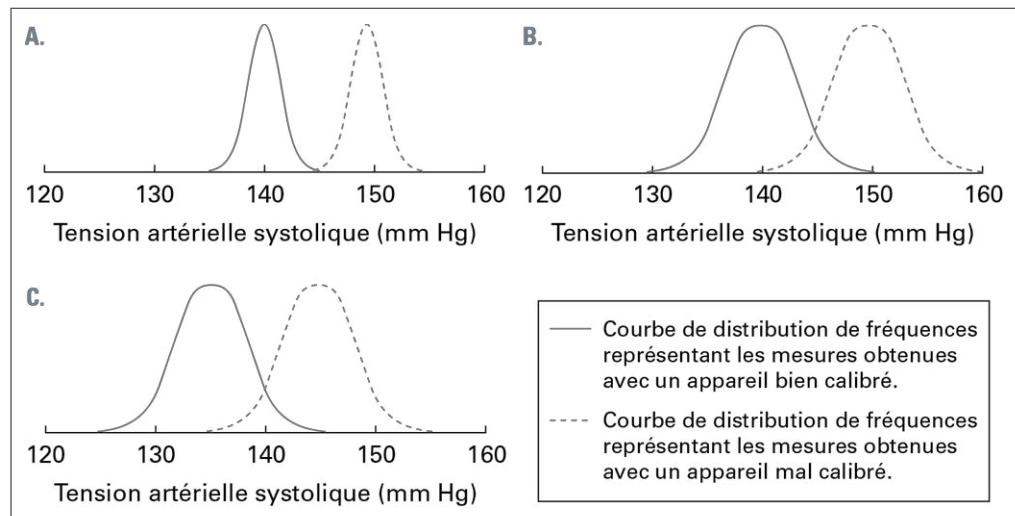
6.1.3 Les relations entre la précision et la validité

Dans une étude, les deux types d'erreurs – systématiques et aléatoires – sont souvent présents.

EXEMPLE 6.2

La figure 6.1, qui représente les mesures de la pression artérielle systolique de Diane, permet de visualiser certaines sources de variations. Les valeurs obtenues au moyen d'un appareil bien calibré (courbes en traits continus) sont en fait les valeurs réelles de la pression artérielle systolique de Diane. En revanche, les valeurs obtenues avec un appareil dont le calibrage est défectueux (courbes en pointillés) sont systématiquement supérieures de 10 mm Hg à la valeur réelle et sont donc biaisées. On trouve en A la plus petite variation, qui s'explique par l'utilisation d'un instrument précis à un moment déterminé de la journée. En B, les variations plus importantes observées par rapport à la valeur réelle de la pression artérielle systolique de Diane, qui était de 140 mm Hg, sont dues au moindre degré de précision de l'appareil. On trouve en C les mesures de la pression artérielle effectuées à différents moments de la journée; l'effet des variations biologiques est ici mis en lumière. La valeur moyenne obtenue en C (135 mm Hg) diffère de la valeur obtenue en A et en B (140 mm Hg). En effet, une valeur observée à un moment donné de la journée, aussi précise soit-elle, ne correspond pas nécessairement à l'ensemble des valeurs observables au cours de cette journée.

FIGURE 6.1 ■ La pression artérielle systolique de Diane



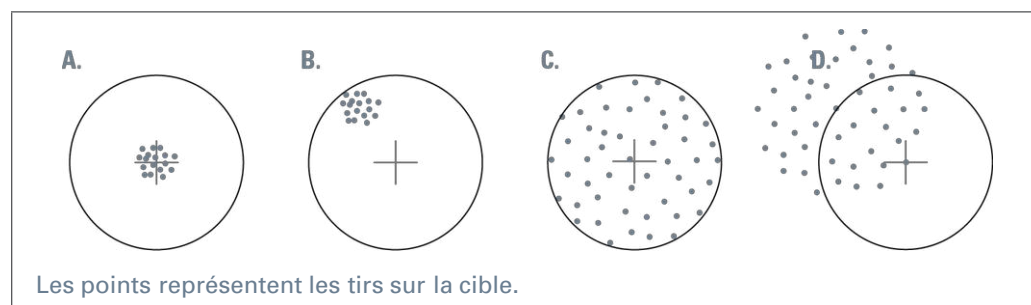
- A. Mesures prises à quelques minutes d'intervalle à l'aide d'un appareil précis
 B. Mesures prises à quelques minutes d'intervalle à l'aide d'un appareil moins précis
 C. Mesures prises au cours de la journée à l'aide d'un appareil précis

L'utilisation d'instruments bien conçus constitue l'un des moyens de réduire l'erreur aléatoire et ainsi d'augmenter la précision de la mesure. Un second moyen consiste à obtenir un plus grand nombre de mesures. En effet, un nombre plus élevé de mesures permet de calculer une moyenne qui résumera plus adéquatement la valeur mesurée. De la même façon, lorsqu'il s'agit d'estimer la pression artérielle systolique moyenne dans la population, l'augmentation de la taille de l'échantillon (le nombre de personnes dont la pression est mesurée) permettra de d'obtenir une mesure plus précise.

La figure 6.2 montre quatre cibles dont le centre est la valeur réelle recherchée: les tirs sur la cible permettent de bien saisir la distinction entre les notions de précision et de validité.

La cible A illustre les mesures obtenues avec un appareil précis et bien calibré (assurant l'obtention de mesures non biaisées); les critères de validité et de précision sont tous deux respectés. La cible B présente des mesures biaisées qui résultent du mauvais calibrage d'un appareil; le critère de validité n'est pas respecté, alors que le critère de précision l'est. La cible C montre les mesures obtenues avec un appareil imprécis, mais bien calibré (assurant l'obtention de mesures non biaisées); le critère de validité est respecté, mais le critère de précision ne l'est pas. Finalement, la cible D correspond à des mesures obtenues avec un appareil à la fois imprécis et mal calibré; les critères de validité et de précision ne sont ni l'un ni l'autre respectés.

FIGURE 6.2 ■ Les notions de précision et de validité en fonction des appareils utilisés



- A. Appareil précis et bien calibré : mesures précises et non biaisées
- B. Appareil précis et mal calibré : mesures précises et biaisées
- C. Appareil imprécis et bien calibré : mesures imprécises et non biaisées
- D. Appareil imprécis et mal calibré : mesures imprécises et biaisées

6.2 Les types de biais

Que les études consultées fournissent des informations sur les valeurs normales des mesures biologiques (glycémie à jeun, pression artérielle, etc.) ou des données sur l'efficacité thérapeutique d'un traitement ou d'un médicament, ou encore qu'elles traitent de l'importance de facteurs de risque, elles sont toutes sujettes à des biais. Trois grands types de biais peuvent influencer sur la validité d'un résultat : les biais de sélection, d'information et de confusion.

6.2.1 Le biais de sélection

Le **biais de sélection** est une erreur systématique qui découle des procédures adoptées pour choisir les sujets à l'étude. Ces procédures peuvent mener à la sélection de sujets qui ne sont pas représentatifs de la population à l'étude ou que l'on ne peut comparer entre eux dans le contexte d'études étiologiques avec groupes de comparaison. Il existe plusieurs types de biais de sélection; ceux dont il est plus particulièrement question ici sont les biais d'échantillonnage ou de survie sélective, d'admission ou de Berkson, de migration ou des perdus de vue, de volontariat, de «bonne santé des travailleurs» et de publication⁴.

Michel est revenu consulter Lisa à la polyclinique et lui a posé une série de questions sur les ITSS et sur les dangers pour la santé qu'entraînent le tabagisme et la consommation d'alcool. Bien que Lisa soit satisfaite de l'information qu'elle lui a livrée, cette rencontre l'incite à approfondir ses connaissances à partir de trois articles récents.

Le biais d'échantillonnage ou de survie sélective

Le premier de ces articles porte sur le lien entre l'alcool et les risques de blessures à l'occasion d'un accident de la route : il s'agit d'une étude cas-témoins. Les cas sont des conducteurs hospitalisés à la suite d'un accident. Les témoins sont des conducteurs impliqués dans un accident, mais qui n'ont pas été blessés. Dans les deux groupes, on mesure la cote de ceux qui dépassent le seuil légal d'alcoolémie, c'est-à-dire la cote d'exposition excessive à l'alcool. Le rapport de ces cotes (RC) est de 3. Il y a donc une association positive entre la consommation excessive d'alcool et le risque de blessures.

Lisa se demande pourtant si cette association ne devrait pas être plus forte. En effet, la population à l'étude est ici l'ensemble des conducteurs accidentés, mais ceux qui sont décédés ne peuvent y participer.

Le tableau 6.2 illustre cette situation en comparant le risque de blessures associé à la consommation excessive d'alcool lorsqu'il est mesuré à partir d'un premier échantillon composé de patients hospitalisés et lorsqu'il est mesuré à partir d'un second échantillon incluant les conducteurs hospitalisés et ceux qui sont décédés avant leur arrivée à l'hôpital. On constate que l'association est effectivement plus forte – le RC de 4,8 étant plus élevé – lorsque l'échantillon est tiré de l'ensemble des conducteurs ayant subi des blessures, graves ou mortelles. Le fait de sélectionner uniquement des patients hospitalisés entraîne donc un biais appelé **biais d'échantillonnage** ou «de survie sélective».

TABLEAU 6.2 ■ Le risque de blessures associé à la consommation excessive d'alcool

A. Patients hospitalisés

	Consommation +	Consommation –
Blessures +	50	40
Blessures –	50	120

RC = 3, soit $(50/40) \div (50/120)$

B. Patients hospitalisés ou décédés

	Consommation +	Consommation –
Blessures +	90	45
Blessures –	50	120

RC = 4,8, soit $(90/45) \div (50/120)$

Le biais d'échantillonnage survient dans une situation où des personnes, membres de la population à l'étude, sont moins susceptibles d'être admises dans une étude à cause d'une caractéristique comme le fait d'être asymptomatique ou atteint d'une maladie mortelle et de courte durée. Une erreur dans l'estimation de la mesure d'association peut en découler et entraîner, selon les circonstances, une surestimation ou une sous-estimation de sa valeur réelle.

Le biais d'admission ou de Berkson

Le deuxième article porte sur le tabagisme et les maladies pulmonaires obstructives chroniques (MPOC). Selon cette étude, il n'y aurait pas d'association entre le tabagisme et l'apparition de MPOC, le RC étant égal à 1. Il s'agit d'une étude cas-témoins réalisée auprès de patients hospitalisés. Cette étude, qui semble bien faite au premier abord, présente toutefois une lacune. En effet, les fumeurs risquent davantage de souffrir de maladies, et par conséquent d'être hospitalisés, que les non-fumeurs. Les témoins hospitalisés sont donc plus susceptibles d'être des fumeurs que les résidents du quartier (non hospitalisés). La population à l'étude n'est pas représentative de la population cible. Cette situation peut avoir pour effet de diminuer ou de faire disparaître l'association entre la maladie étudiée et le tabagisme. Le tableau 6.3 illustre ce type d'effet.

TABLEAU 6.3 ■ Le tabagisme et les maladies pulmonaires obstructives chroniques (MPOC)

A. Patients hospitalisés

	Tabagisme +	Tabagisme -
MPOC	50	100
Autres causes	50	100

RC = 1, soit $(50/100) \div (50/100)$

B. Résidents du quartier (non hospitalisés)

	Tabagisme +	Tabagisme -
MPOC	50	100
Autres causes	25	125

RC = 2,5, soit $(50/100) \div (25/125)$

Ce type de biais est appelé **biais d'admission** ou **biais de Berkson**. Il apparaît lorsque les personnes sélectionnées dans la population à l'étude sont différentes de la population cible à cause de l'exposition étudiée. Il peut en résulter une surestimation ou une sous-estimation de l'association.

Le biais de migration ou des perdus de vue

Le troisième article porte sur un essai clinique au cours duquel on a comparé un antibiotique récemment mis au point au médicament habituellement prescrit pour le traitement de la chlamydia. Tel que présenté dans le tableau 6.4, Lisa apprend que 94 % des patients ayant reçu le nouveau médicament sont guéris, comparativement à 86 % de ceux qui ont suivi le traitement habituel. Elle remarque toutefois que 30 % des patients du groupe traité avec le nouvel antibiotique ne se sont pas présentés pour l'évaluation de l'efficacité du traitement, proportion qui, par contre, n'est que de 5 % dans le groupe ayant reçu le traitement courant. Lisa s'interroge : si les résultats de

l'antibiothérapie chez les patients perdus de vue en cours d'étude différaient substantiellement de ceux constatés chez les patients qui ont été observés pour la durée complète de l'étude, il serait possible que l'efficacité du médicament soit surévaluée ou sous-évaluée, cet antibiotique pouvant avoir été complètement inefficace pour la totalité des patients perdus de vue. Seule la connaissance des effets du traitement chez les patients qui ne se sont pas présentés à l'évaluation permettrait d'avoir une idée juste de l'efficacité réelle du nouveau médicament.

Le tableau 6.4B, qui est constitué en supposant que les patients perdus de vue en cours d'étude ont pu être rejoints pour une évaluation de l'efficacité du traitement, illustre cette situation. Cinquante pour cent de ces patients utilisant le nouveau traitement, soit 15 patients sur 30, sont guéris, alors que 20 % des patients soumis au traitement habituel et qu'on avait perdus de vue, soit 1 patient sur 5, le sont aussi. En tenant compte des perdus de vue des deux groupes, on arrive au résultat suivant : aux 66 patients guéris par le nouveau médicament s'ajoutent les 15 patients retrouvés guéris, ce qui donne un total de 81 guéris sur 100 ; aux 82 patients guéris par le traitement habituel s'ajoute une personne retrouvée guérie, ce qui donne un total de 83 guéris sur 100. Dans cette situation, le traitement récemment mis au point ne semble donc pas plus efficace que le traitement courant.

TABLEAU 6.4 ■ L'efficacité d'un traitement nouveau comparée à celle du traitement habituel

A. Sans les perdus de vue

	Nouveau traitement	Traitement habituel
Patients guéris	66 (94%)	82 (86%)
Patients rejoints	70	95

B. Avec les perdus de vue

	Nouveau traitement	Traitement habituel
Patients guéris	81 (81%)	83 (83%)
Patients rejoints	100	100

Le type de biais décrit dans cet exemple est dit **biais de migration**, aussi appelé des perdus de vue, en référence au départ des participants. Le tableau 6.4 illustre qu'il est possible de présenter les résultats d'un essai clinique en tenant compte des participants perdus de vue. Cet exemple démontre l'importance de préciser dans un article la façon dont sont analysés les sujets perdus en cours de suivi. On constate que les résultats et la conclusion diffèrent grandement selon que l'on inclut ou non ces sujets. Il aurait ainsi été souhaitable que l'article consulté par Lisa parle davantage des répercussions possibles de ce biais sur les résultats.

Le biais de volontariat

Le **biais de volontariat** découle de la stratégie de recrutement de l'échantillon des sujets à l'étude. Il réside dans le fait que les caractéristiques des sujets volontaires (âge, sexe, état de santé, etc.) participant à une étude ne sont pas représentatives de l'ensemble des caractéristiques des individus qui composent la population à l'étude. Ces caractéristiques particulières peuvent influencer sur la mesure d'association.

EXEMPLE 6.3

Des chercheurs doivent mesurer la capacité auditive des personnes qui n'ont jamais eu d'atteintes auditives et qui n'ont pas été exposées à des facteurs de risque susceptibles de détériorer leur ouïe, dans le but d'établir des valeurs de référence. Pour recruter leur échantillon, ils diffusent des annonces dans les médias locaux et sociaux, invitant les gens à passer un examen auditif sur une base volontaire. Il est alors fort probable que les personnes les plus intéressées par l'examen soupçonnent déjà la présence d'un problème d'audition et souhaitent profiter de l'occasion pour en faire l'évaluation. Dans ce contexte, l'échantillon ne sera pas représentatif de la population à l'étude. On y trouvera en plus grande proportion des individus ayant un problème auditif. Ce biais de volontariat aura ici pour effet de sous-estimer la capacité auditive de la population à l'étude.

Le biais de « bonne santé des travailleurs »

Le **biais de « bonne santé des travailleurs »**, traduction littérale de *healthy worker effect*, tient au fait que des groupes de travailleurs sélectionnés dans une étude présentent un taux de morbidité ou de mortalité plus faible que celui de la population en général. Ce phénomène pour le moins étonnant s'explique principalement par le fait que les personnes en bonne santé sont plus susceptibles de travailler et de conserver un emploi que les personnes aux prises avec un ou des problèmes de santé faisant partie de la population en général⁵. Une recherche réalisée dans une population à l'étude de travailleurs ne pourra donc être généralisée que dans une population cible de travailleurs.

Le biais de publication

Un autre biais d'importance grandissante est le **biais de publication**. Ce biais de sélection est lié à la tendance qui consiste à publier principalement les articles qui appuient les hypothèses de recherche établies. Ce biais est particulièrement notable dans les essais cliniques et dans les études sur les risques environnementaux. Un biais de publication est présent lorsque les auteurs d'une étude ne réussissent pas à faire publier leurs résultats dits négatifs (mesure d'association nulle ou allant dans le sens contraire de l'hypothèse de recherche). Le biais de publication touche particulièrement les articles de **synthèse (qualitative ou quantitative)** lorsque ceux-ci ne portent que sur des études publiées. À cet égard, le chercheur doit tenter de trouver des études non publiées et de les inclure dans son article de synthèse, ce qui n'est pas toujours évident.

EXEMPLE 6.4

Dans un article du *New England Journal of Medicine*⁶, on a démontré l'importance du biais de publication dans l'évaluation de l'efficacité des antidépresseurs en comparant les résultats des études enregistrées à la Food and Drug Administration (FDA) à ceux qui ont ultimement été publiés dans les journaux scientifiques. Sur 74 études enregistrées à la FDA, 31 % d'entre elles n'ont pas été publiées. Trente-sept études aux résultats jugés positifs selon la FDA ont été publiées, alors qu'une seule étude dans le même cas ne l'a pas été. Des 37 études restantes dont les

EXEMPLE 6.4 (suite)

résultats étaient négatifs ou douteux, 22 n'ont pas été publiées et 11 l'ont été de manière à ce que les résultats apparaissent positifs. Selon la littérature publiée, il appert que 94 % des essais menés présentaient des résultats positifs, tandis que, selon la FDA, 51 % d'entre eux l'étaient réellement.

La réduction des effets du biais de sélection

Il est essentiel de se montrer attentif à la façon dont a été effectuée la sélection des participants à une étude, car il peut en résulter des biais qui influenceront sur les conclusions de façon importante. Il faut faire preuve d'esprit critique devant les études consultées si l'on veut en retirer une information utile à la pratique clinique. Or, la première attitude à adopter est de voir si les auteurs ont cherché à éliminer ou à réduire les biais de sélection. Ainsi, il importe de se demander si les participants à l'étude ont été choisis par tirage au sort, puisque la sélection aléatoire constitue la meilleure façon d'obtenir un échantillon représentatif de la population à l'étude. Cette façon de procéder permet d'éliminer le biais de volontariat.

Les chercheurs doivent préciser les critères d'inclusion et d'exclusion des sujets de leur recherche, ce qui permet, par comparaison avec la population à l'étude (lorsque l'on dispose de données sur cette population), de juger de la représentativité de l'échantillon.

À retenir**Les biais de sélection**

Type de biais	Prévention du biais	Contrôle du biais au moment de l'analyse
Échantillonnage ou survie sélective	<ul style="list-style-type: none"> S'assurer que toutes les personnes de la population à l'étude ont une chance égale d'y participer. Utiliser un échantillonnage aléatoire stratifié. 	Détermination de son influence sur la mesure
Admission ou Berkson	<ul style="list-style-type: none"> S'assurer que la population à l'étude est représentative de la population cible. 	Détermination de son influence sur la mesure
Migration ou perdus de vue	<ul style="list-style-type: none"> Utiliser des stratégies pour minimiser les perdus de vue. 	Analyse de sensibilité
Volontariat	<ul style="list-style-type: none"> Utiliser un échantillonnage aléatoire. 	Détermination de son influence sur la mesure
« Bonne santé des travailleurs »	<ul style="list-style-type: none"> Éviter de comparer un groupe de travailleurs exposés à un facteur de risque au travail à un groupe issu de la population en général. 	Détermination de son influence sur la mesure
Publication	<ul style="list-style-type: none"> Prendre des mesures réglementaires pour créer un registre international des études entreprises sans égard à la publication des résultats. 	Recherche systématique de toutes les études non publiées sur le sujet

Note : Tous ces biais peuvent entraîner une surestimation ou une sous-estimation de la mesure d'intérêt.

6.2.2 Le biais d'information

Le **biais d'information** est une erreur systématique entraînée par la mesure ou l'observation incorrecte de l'exposition ou de la maladie. En contexte clinique, des biais d'information peuvent survenir à toutes les étapes du processus diagnostique : à l'anamnèse, au cours de l'examen clinique ou des examens paracliniques, etc. À la lecture d'articles portant sur des études, l'attention portée à la qualité des instruments de mesure utilisés et à la façon dont les chercheurs ont procédé pour colliger les informations renseigne le lecteur sur la présence ou l'absence de biais d'information.

Michel, lors de sa première visite, présentait des symptômes que Lisa pouvait associer à une ITSS : douleur (sensation de brûlure) au moment d'uriner, apparition de sécrétions claires ou purulentes ou encore mictions plus fréquentes. Lisa a donc procédé à un prélèvement urétral ; les résultats des examens de laboratoire appropriés (elle suspecte une gonococcie ou une chlamydia) se révèlent négatifs. Que s'est-il passé ? Michel présentait peut-être un autre problème de santé qui n'a pas été envisagé, par exemple, une urétrite causée par un autre agent infectieux. Il est également possible que Michel soit effectivement atteint d'une infection à gonocoque ou à *Chlamydia*, mais qu'elle n'ait pu être mise en évidence en raison des lacunes que comportent les analyses de laboratoire demandées. Ces examens peuvent donc entraîner une erreur de classification : les résultats pourraient indiquer une absence du microorganisme alors que Michel est bel et bien infecté.

EXEMPLE 6.5

Ce matin, à son arrivée à la polyclinique, Lisa prend connaissance d'un avis provenant du centre de santé publique dans lequel est annoncée une recrudescence des cas de coqueluche dans la région. On demande d'être vigilant quand des personnes se présentent en consultation avec des symptômes pouvant être associés à cette maladie contagieuse. On rappelle que cette infection peut être traitée et qu'il est possible d'en réduire la propagation. En outre, on joint à cet avis les résultats d'une enquête récente montrant qu'un pourcentage élevé de médecins ne cherchent pas systématiquement à déceler les cas de coqueluche. L'enquête précise certaines des circonstances où l'on n'envisage généralement pas un tel diagnostic : personnes ayant reçu une vaccination complète, adolescents ou adultes, personnes ne présentant pas les symptômes habituels, etc.

L'exemple précédent montre que le biais d'information n'est pas seulement lié à l'utilisation d'instruments de mesure imparfaits, mais qu'il peut aussi être entraîné par les pratiques de l'observateur. Si le médecin ne soupçonne pas la présence d'un problème de santé ou ne pose pas les gestes permettant de confirmer ou d'infirmer une hypothèse diagnostique, il y a biais d'information : la personne peut être atteinte de l'affection, la coqueluche par exemple, sans que l'on dépiste la maladie.

Lisa a récemment pris connaissance de deux études cas-témoins portant sur la relation entre la consommation d'alcool et l'apparition de cancers. Ces études permettent

d'illustrer les deux types d'erreurs de classification associés au biais d'information : l'erreur non différentielle et l'erreur différentielle.

L'erreur de classification non différentielle

Supposons que la mesure réelle de l'association entre la quantité déterminée d'alcool consommé et le cancer corresponde à un RC de 1,5. Dans la première étude, on obtient l'information sur la quantité d'alcool consommé à l'aide d'un questionnaire rempli au cours d'une entrevue avec le participant, qu'il soit atteint (cas) ou non atteint (témoin) d'un cancer. Le fait de classer les individus dans les catégories « exposés » et « non exposés » en utilisant un même instrument imparfait entraîne une **erreur de classification non différentielle**. On l'appelle « non différentielle » parce qu'elle affecte de la même façon les cas et les témoins. Ce type d'erreur a pour effet de mener à une sous-estimation de la vraie mesure d'association, qui tendra vers la valeur nulle. Dans cet exemple, le RC résultant de l'association pourrait se situer entre 1,0 (la valeur nulle pour un RC) et 1,5. Dans une étude cas-témoins, il y a erreur de classification non différentielle si l'erreur qui influe sur la mesure de l'exposition est la même pour les cas et les témoins. Dans une étude de cohorte, il y a erreur de classification non différentielle si l'erreur qui influe sur la mesure de la maladie est la même pour les groupes exposés et non exposés.

L'erreur de classification différentielle

Dans la seconde étude, on a obtenu l'information sur la quantité d'alcool consommé par les patients atteints d'un cancer au cours d'une entrevue, alors que pour les témoins, on a procédé par questionnaire autoadministré. Dans ce contexte, des instruments de mesure différents ont engendré une **erreur de classification différentielle**. On la qualifie ainsi parce qu'elle affecte différemment la mesure de l'exposition selon les cas et les témoins. Une telle erreur, due à l'utilisation d'instruments de mesure différents ou au fait que les observateurs ne procèdent pas de la même façon, peut conduire à une surestimation ou à une sous-estimation de la mesure d'association réelle. Toujours dans cet exemple, le RC résultant de l'association entre une quantité déterminée d'alcool et le cancer pourrait être inférieur ou supérieur à 1,5.

Dans une étude cas-témoins, il y a erreur de classification différentielle si l'erreur qui influe sur la mesure de l'exposition n'est pas la même pour les cas et les témoins. Dans une étude de cohorte, il y a erreur de classification différentielle si l'erreur qui influe sur la mesure de la maladie n'est pas la même pour les groupes exposés et non exposés.

Le biais de rappel

Enfin, une autre erreur différentielle importante peut survenir dans les études cas-témoins et transversales : il s'agit du **biais de rappel**. Les personnes présentant un problème de santé peuvent avoir tendance à se rappeler et à déclarer les expositions antérieures à un facteur quelconque de façon différente que ne le feraient celles qui sont en bonne santé.

EXEMPLE 6.6

Une personne atteinte d'un cancer peut se rappeler de façon plus détaillée son exposition à des solvants au travail si elle estime qu'il peut exister un lien entre cette exposition et l'apparition du cancer. Ce biais conduit ici à une surestimation de l'association entre le facteur et la maladie en question. À l'inverse, une femme ayant eu un bébé né avec des malformations pourrait minimiser ses expositions plus à risque (par exemple, son tabagisme, sa consommation d'alcool ou de stupéfiants, etc.), ce qui entraînerait ici une sous-estimation de l'association entre l'exposition et la maladie en question.

La réduction des effets du biais d'information

Comment évite-t-on les biais d'information? La formation en sciences de la santé doit permettre d'acquérir les compétences requises dans un contexte clinique pour effectuer adéquatement l'anamnèse et l'examen physique, bien évaluer la pertinence des examens paracliniques, etc. Le but est simple : être en mesure d'obtenir de l'information valide susceptible d'améliorer l'état de santé du patient et de réduire le plus possible le nombre d'interventions inutiles ainsi que les risques d'effets indésirables qu'elles comportent.

Les réponses à ces questions permettront de bien évaluer l'effort déployé par les auteurs des études pour réduire les biais d'information ou pour en connaître les effets.

- La méthode ayant permis d'obtenir les informations est-elle décrite adéquatement?
- Les auteurs mentionnent-ils les caractéristiques des instruments utilisés ainsi que leurs limites?
- S'il y a plus d'un observateur pour la collecte des données, ont-ils reçu une formation homogène, et leur travail est-il supervisé? Y a-t-il des contrôles de qualité?
- Dans les essais cliniques⁷, s'est-on assuré : 1. que les sujets ignorent le traitement qu'ils reçoivent (simple aveugle ou insu)? 2. que les sujets et les chercheurs qui évaluent les résultats ignorent quel traitement est administré (double aveugle ou insu)? et 3. qu'en plus, le biostatisticien qui évalue les résultats ignore aussi quel traitement est administré (triple aveugle ou insu)?
- Dans la discussion d'un article, tient-on compte de l'influence potentielle de ces biais sur les résultats obtenus? A-t-on demandé directement aux sujets s'ils se doutaient du traitement qu'ils recevaient?

À retenir**Le biais d'information selon le type d'étude**

Type d'étude	Source du biais	Prévention du biais	Contrôle du biais au moment de l'analyse
Descriptive	Erreur systématique liée à la mesure de l'exposition ou de la maladie	• Instruments de mesure valides	Détermination de son influence sur la mesure →

Le biais d'information selon le type d'étude (*suite*)

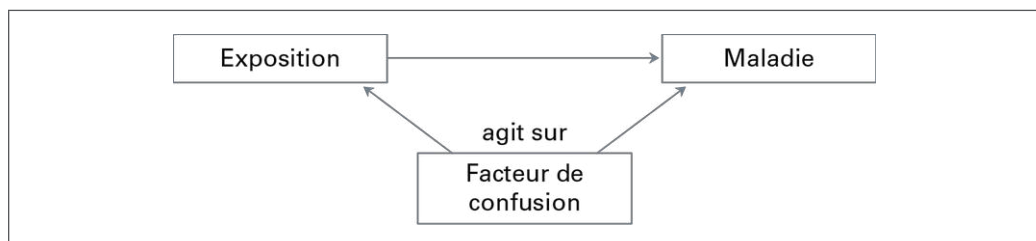
Type d'étude	Source du biais	Prévention du biais	Contrôle du biais au moment de l'analyse
À visée étiologique	Erreur systématique liée à la mesure de l'exposition (dans les études cas-témoins) ou de la maladie (dans les études non expérimentales), affectant différemment les groupes de comparaison	<ul style="list-style-type: none"> • Instruments de mesure valides 	Détermination de son influence sur la mesure
Expérimentale	Erreur systématique liée à la mesure de la maladie, affectant différemment les groupes de comparaison	<ul style="list-style-type: none"> • Instruments de mesure valides • Technique en simple, double ou triple aveugle ou insu 	Vérification du respect de l'aveugle

Note: Tous ces biais peuvent entraîner une surestimation ou une sous-estimation de la mesure d'intérêt.

6.2.3 Le biais de confusion

Le **biais de confusion** désigne l'erreur systématique induite par la présence d'un facteur associé de façon indépendante tant à la maladie qu'au facteur d'exposition à l'étude. Ce facteur ne doit pas appartenir à la chaîne causale de la maladie, c'est-à-dire que le facteur de confusion est un facteur de risque ou une cause de la maladie, mais n'est pas un facteur de risque ou une cause de l'exposition. La présence du facteur de confusion peut résulter des caractéristiques de la population à l'étude ou de la procédure de sélection des sujets. La figure 6.3 illustre la relation entre une exposition, une maladie et un facteur de confusion.

FIGURE 6.3 ■ La relation entre une exposition, une maladie et un facteur de confusion



Pour illustrer le concept de chaîne causale, on peut considérer les deux exemples suivants.

EXEMPLE 6.7

On s'interroge sur l'effet de l'âge dans la relation entre le tabagisme et le cancer du poumon. L'âge est associé à la fois au tabagisme (plus on est âgé, plus la durée d'exposition au tabac est susceptible d'être élevée) et au cancer du poumon (l'augmentation du nombre d'années de vie accroît le risque d'être atteint). Par ailleurs, l'âge ne constitue pas une cause du tabagisme. Conséquemment, l'âge est un facteur de confusion à considérer dans la relation entre le tabagisme et le cancer du poumon.

EXEMPLE 6.8

On se questionne sur l'effet de la consommation de lipides dans la relation entre la cholestérolémie et les maladies cardiovasculaires (MCV). La consommation lipidique est associée à la fois à la cholestérolémie et aux MCV. Or, la consommation de lipides influe sur la cholestérolémie qui, elle, a une incidence sur les MCV; la cholestérolémie se trouve dans la chaîne causale partant de la consommation alimentaire et menant aux MCV. Conséquemment, la consommation de lipides n'est pas un facteur de confusion dans la relation entre la cholestérolémie et les MCV.

Lisa analyse maintenant les résultats d'une étude de cohorte portant sur le risque pour les participants d'être infectés par le virus de l'immunodéficience humaine (VIH) selon qu'ils font usage (UDI) ou non (non-UDI) de drogues intraveineuses⁸; les auteurs y ont pris en compte les relations sexuelles avec plus d'un partenaire. Le tableau 6.5 présente dans l'ordre le risque relatif brut (RR_b) pour l'ensemble des personnes sans égard à l'activité sexuelle ainsi que, sous forme de strates (ou sous-groupes), les résultats des risques relatifs spécifiques (RR_s) pour les personnes ayant un partenaire sexuel stable (strate A) et pour celles ayant plus d'un partenaire sexuel (strate B).

La présence de plus d'un partenaire sexuel stable pourrait être ici un facteur de confusion. Pour induire un biais, cette variable doit être associée de façon indépendante au facteur d'exposition et à la maladie sans être une cause du facteur d'exposition ni un facteur intermédiaire entre l'exposition et la maladie.

En premier lieu, Lisa vérifie s'il y a, dans les données, une association entre la présence de plus d'un partenaire sexuel et le facteur d'exposition (UDI ou non-UDI). Pour ce faire, elle évalue si la proportion des participants ayant des partenaires sexuels multiples est différente selon qu'ils appartiennent au groupe des UDI ou au groupe des non-UDI. On compte dans l'étude 1 500 UDI, dont 1 000 ont plus d'un partenaire; la proportion des UDI ayant des partenaires multiples est donc de 0,66. Par ailleurs, on y dénombre 2 500 non-UDI, dont 500 ont des partenaires multiples; la proportion des non-UDI ayant des partenaires multiples est donc de 0,20. Il existe une association évidente entre le facteur de confusion présumé et le facteur d'exposition. Par ailleurs, les partenaires multiples ne sont pas une cause connue d'utilisation de drogues intraveineuses.

En second lieu, Lisa tente d'établir s'il y a, dans les données, une association entre la présence de plus d'un partenaire sexuel et celle de la maladie (séroconversion secondaire à une infection par le VIH), puisque le fait d'avoir des partenaires sexuels multiples accroît le risque de séroconversion dans la population. Comment cela se reflète-t-il dans les données? Pour le découvrir, il faut mesurer les fluctuations du risque selon que les participants ont ou non des partenaires multiples. Le risque de séroconversion encouru par les non-UDI est de 20/2 000 (0,01) chez les participants ayant un seul partenaire; il s'élève à 10/500 (0,02) chez les participants ayant des partenaires multiples. Dans le groupe des non-UDI, le risque de séroconversion est donc deux fois plus grand pour les personnes qui ont plusieurs partenaires que pour les autres (0,02/0,01 = 2). Par conséquent, la présence de plus d'un partenaire sexuel constitue un facteur de risque pour la maladie. Sur cette base, Lisa peut donc affirmer que cette variable est bel et bien un facteur de confusion.

TABLEAU 6.5 ■ Le risque de séroconversion (VIH): étude de cohorte

Ensemble des personnes, sans égard au nombre de partenaires

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	250	30	280
VIH -	1 250	2 470	3 720
Total	1 500	2 500	4 000

$$RR_b = 13,9, \text{ soit } (250/1\ 500) \div (30/2\ 500)$$

Strate A: Personnes ayant un partenaire sexuel stable

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	50	20	70
VIH -	450	1 980	2 430
Total	500	2 000	2 500

$$RR_s = 10, \text{ soit } (50/500) \div (20/2\ 000)$$

Strate B: Personnes ayant plus d'un partenaire sexuel

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	200	10	210
VIH -	800	490	1 290
Total	1 000	500	1 500

$$RR_s = 10, \text{ soit } (200/1\ 000) \div (10/500)$$

Comment cette confusion se manifeste-t-elle? En examinant les résultats du tableau, Lisa note que les UDI sont 10 fois plus susceptibles de devenir séropositifs que les non-UDI, qu'ils aient ou non plusieurs partenaires sexuels (RR_s de la strate A et de la strate B). Toutefois, la mesure d'association brute obtenue lorsque l'on ne tient pas compte du nombre de partenaires ($RR_b = 13,9$) est différente des mesures spécifiques, indiquant des RR_s de 10. En général, lorsque la mesure brute n'est pas comprise à l'intérieur de l'intervalle déterminé par la mesure spécifique la plus élevée

et la mesure spécifique la moins élevée (les mesures spécifiques ne sont pas toujours égales), c'est qu'il y a présence d'un biais de confusion.

Le contrôle du biais de confusion

On peut contrôler l'effet d'un facteur de confusion en étudiant des strates (dans le tableau 6.5, il s'agit des strates A et B) dans lesquelles ce facteur est absent (strate A) ou présent (strate B) et en utilisant les mesures propres à chacune de ces strates plutôt que la mesure brute. D'autres moyens pour contrôler ou réduire les biais de confusion sont utilisés, selon le cas, au moment de l'échantillonnage ou de l'analyse.

Au cours de l'échantillonnage, on peut procéder par **restriction** ou par **assortiment**. Si l'on procède par restriction, on exclut de l'étude les gens qui sont exposés au facteur de risque potentiellement confondant. Dans l'exemple précédent, on pourrait exclure de l'étude les individus qui ont plus d'un partenaire sexuel, qu'ils fassent ou non usage de drogues intraveineuses.

Si l'on procède par assortiment, on rendra les groupes étudiés comparables au regard du facteur à contrôler. Ainsi, toujours dans l'exemple précédent, on pourrait s'assurer que la proportion des personnes ayant plus d'un partenaire sexuel chez l'ensemble des individus soit la même dans le groupe des UDI et dans celui des non-UDI. Dans le tableau 6.5, Lisa constate que les deux tiers des UDI, soit 1 000/1 500, ont plus d'un partenaire sexuel, alors que cette proportion est d'un cinquième, soit 500/2 500, dans le groupe des non-UDI. Pour éliminer le biais de confusion, il faudrait donc établir une proportion équivalente de non-UDI ayant plus d'un partenaire, soit les deux tiers, en recrutant par exemple 2 000 non-UDI ayant plus d'un partenaire et 1 000 UDI en ayant un seul. Le tableau 6.6 présente cette nouvelle situation. Le calcul du RR_b donne maintenant la même valeur que celles observées sur les strates A et B, ce qui signifie qu'il y a absence de biais de confusion.

TABLEAU 6.6 ■ Le risque de séroconversion (VIH) : étude sans biais de confusion

Ensemble des personnes, sans égard au nombre de partenaires

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	250	50	300
VIH -	1 250	2 950	4 200
Total	1 500	3 000	4 500

$RR_b = 10$, soit $(250/1 500) \div (50/3 000)$

Strate A: Personnes ayant un partenaire sexuel stable

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	50	10	60
VIH -	450	990	1 440
Total	500	1 000	1 500

$RR_s = 10$, soit $(50/500) \div (10/1 000)$



TABLEAU 6.6 ■ Le risque de séroconversion (VIH) : étude sans biais de confusion (suite)**Strate B: Personnes ayant plus d'un partenaire sexuel**

	Drogues intraveineuses +	Drogues intraveineuses -	Total
VIH +	200	40	240
VIH -	800	1 960	2 760
Total	1 000	2 000	3 000

$$RR_s = 10, \text{ soit } (200/1\ 000) \div (40/2\ 000)$$

Dans le cadre de l'essai clinique, la répartition des sujets dans des groupes de comparaison en procédant par tirage au sort (hasardisation) permet de prévenir un biais de confusion. Il est toutefois nécessaire de vérifier que cette distribution au hasard a bien fonctionné. Ainsi, dans un article rapportant les résultats d'un essai clinique, on présente habituellement les caractéristiques des participants de chaque groupe (répartition selon l'âge, facteurs de risque, niveau d'éducation, ethnie, etc.) afin de s'assurer que les groupes sont effectivement comparables. S'il existe entre les groupes des différences qui réduisent la comparabilité, elles pourraient être prises en compte au moment de l'analyse des résultats, car ces différences peuvent introduire un biais de confusion.

Le contrôle des biais de confusion peut se faire au cours de l'analyse, soit par **stratification**, soit par l'utilisation de modèles mathématiques. Les tableaux 6.5 et 6.6 exposent la méthode de contrôle par stratification; celle-ci consiste à calculer des mesures d'association spécifiques en fonction de la présence ou de l'absence du facteur de confusion. Il est également possible de contrôler le biais de confusion grâce à l'ajustement de la mesure brute, en ayant recours à des méthodes d'ajustement⁹ ou à des méthodes de régression que nous aborderons dans le chapitre 12.

À retenir

Le biais de confusion selon le type d'étude

Type d'étude	Source du biais	Prévention du biais	Contrôle du biais au moment de l'analyse statistique
Descriptive et à visée étiologique	Erreur systématique liée à l'effet d'un tiers facteur dans la relation entre l'exposition et la maladie	Détermination de toutes les variables potentiellement confondantes et prise en compte de ces variables dans l'analyse	<ul style="list-style-type: none"> • Appariement ou restriction au cours de l'échantillonnage • Analyse stratifiée • Analyse ajustée
Expérimentale		Hasardisation des sujets	<ul style="list-style-type: none"> • Analyse en intention de traitement (<i>voir le chapitre 13</i>) • Analyse ajustée

Note: Tous ces biais peuvent entraîner une surestimation ou une sous-estimation de la mesure d'intérêt.

En clinique, le fait de connaître l'existence de biais est très utile, car cela permet de s'interroger sur la qualité de l'anamnèse – et donc sur la façon la plus adéquate de mener l'entrevue –, sur l'utilisation appropriée des instruments, ou encore sur la pertinence de recourir à certains examens paracliniques. Aussi cherchera-t-on, à la lecture d'articles scientifiques, à savoir si les auteurs ont tenu compte des biais potentiels et utilisé une ou des méthodes de contrôle. Dans les articles de qualité, les auteurs ne craignent pas de discuter de l'ensemble des biais pouvant entacher les résultats de leur étude ; le lecteur peut donc connaître les limites des méthodes utilisées.

Enfin, plusieurs autres biais spécifiques n'ont pas été discutés dans ce chapitre (biais de désirabilité sociale, de détection, de performance, de confusion par indication, etc.). Retenez qu'ils sont tous assignés à l'un des trois grands types de biais.

En résumé

La notion de justesse, soit la validité et la précision des mesures en épidémiologie, permet de juger de la validité et de l'utilité d'une publication scientifique. La validité fait référence à la capacité d'appréhender correctement un phénomène ou une mesure. Il existe trois grands types de biais pouvant affecter la

validité d'une mesure : les biais de sélection, d'information et de confusion. La précision tient à l'absence d'erreurs aléatoires ou dues au hasard. Il importe d'accorder une attention particulière à ces notions pour juger de la qualité des études et de leur applicabilité à la pratique professionnelle.

Exercices



Voir la section complète des exercices sur le Web.

1. Vous réalisez une étude visant à mesurer la capacité auditive normale des individus adultes dans une population. Les personnes sont invitées à passer un test d'audition sur une base volontaire. Quel type de biais pourra influencer sur les résultats de l'étude ?
 - a) Biais de sélection.
 - b) Biais d'information.
 - c) Biais de confusion.
2. Quelle est la meilleure stratégie pour éviter un biais de sélection ?
 - a) Établir des critères précis sur les caractéristiques des individus à sélectionner.
 - b) Sélectionner les individus au hasard.
 - c) Augmenter le nombre d'individus dans l'échantillon observé.
3. Dans une étude à visée étiologique de cohorte, on veut vérifier la relation entre la capacité pulmonaire et l'exposition aux poussières minérales chez les travailleurs des mines d'amiante et de silice. La capacité pulmonaire des individus exposés est évaluée à l'aide d'un spiromètre bien calibré alors que la capacité pulmonaire des personnes non exposées est évaluée à l'aide d'un spiromètre mal calibré qui sous-estime leur capacité pulmonaire réelle. Ce biais d'information entraînera-t-il une erreur différentielle ou une erreur non différentielle ?
4. Quelle sera l'effet de ce biais d'information sur la mesure d'association à l'énoncé de la question 3 ?
 - a) Sous-estimation du RR.
 - b) Surestimation du RR.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 7

Les examens diagnostiques

Collaboration à la 3^e édition :

Marie-Hélène Mayrand

Helen Trottier

Thomas Matukala Nkosi

Antoine Lutumba Ntetu

Au terme de ce chapitre, vous pourrez :

- juger de la validité et de la pertinence d'un test diagnostique.

Objectifs spécifiques	Contenu
7.1 Classifier les résultats d'examens diagnostiques et de référence à l'aide d'un tableau de contingence.	7.1 La classification des résultats d'un examen diagnostique 7.1.1 La comparaison d'un examen diagnostique à un examen de référence 7.1.2 La classification des résultats
7.2 Définir les notions de base permettant d'apprécier la validité interne d'un examen diagnostique.	7.2 La validité intrinsèque: sensibilité (SE) et spécificité (SP) 7.2.1 Le calcul de la sensibilité et de la spécificité 7.2.2 Le choix d'un examen sensible ou spécifique 7.2.3 L'influence du seuil de positivité
7.3 Définir les notions de base permettant d'apprécier la validité prédictive d'un examen diagnostique.	7.3 La validité extrinsèque: les valeurs prédictives positives et négatives 7.3.1 Les valeurs prédictives positives et négatives 7.3.2 Le calcul des valeurs prédictives 7.3.3 L'influence de la prévalence et de la validité intrinsèque
7.4 Déterminer les stratégies d'investigation à l'aide des examens diagnostiques ainsi que les paramètres cliniques qui influenceront le choix des tests.	7.4 Les combinaisons d'examens 7.4.1 La combinaison d'examens en parallèle 7.4.2 La combinaison d'examens en série
7.5 Déterminer les sources de variations susceptibles d'influencer le résultat d'un examen diagnostique.	7.5 Les sources de variation dans les examens diagnostiques 7.5.1 La mesure de la variation: la mesure d'accord Kappa (K) 7.5.2 Les variations attribuables à l'examineur 7.5.3 Les variations attribuables à la personne examinée 7.5.4 Les variations attribuables à l'examen



Mise en situation

Il est 20 h lorsque Paule consulte l'omnipraticienne Lisa au service de consultation sans rendez-vous de la polyclinique Milo. Ces derniers temps, Paule est fatiguée, se sent facilement essoufflée et présente une toux sèche. Depuis 24 heures, elle ressent une douleur à la poitrine, à droite, qui est très inconfortable et qui augmente lorsqu'elle inspire. Paule est âgée de 42 ans; elle est mariée et a deux enfants. Le plus vieux a trois ans, fréquente une garderie et se remet d'un gros rhume. Le plus jeune a un mois. Elle allaite son bébé et s'est bien remise de l'accouchement, bien qu'elle ait dû subir une césarienne urgente. Elle ne formule aucune autre plainte.

Paule a toujours eu une très bonne santé. Elle ne fume pas et ne prend aucun médicament. Elle est fille unique; son père est décédé à 72 ans d'un infarctus du myocarde alors que sa mère, âgée de 65 ans, demeure en bonne santé.

À l'examen, Lisa constate que Paule présente tout de même un bon état général. Sa température est à 38 °C. Sa pression artérielle et son rythme respiratoire se situent dans les limites normales, mais son pouls est un peu rapide à 90 battements par minute. L'examen du cœur, des poumons et des membres inférieurs se révèle normal. La radiographie pulmonaire est également normale.

Lisa explique à sa patiente qu'étant donné le site et les caractéristiques de la douleur qu'elle

ressent, elle ne peut exclure la possibilité d'une embolie pulmonaire. Elle suggère à Paule de subir une prise de sang afin de procéder à un dosage des d-dimères, un test de première ligne recommandé quand on entretient la possibilité d'un diagnostic d'embolie pulmonaire. Les d-dimères sont un produit de la dégradation de la fibrine, qui entre dans la constitution des caillots sanguins.

Pertinence clinique

Lisa est formée pour établir un diagnostic à partir des combinaisons de signes et des symptômes que manifeste un patient. Ceux que présente Paule sont peut-être attribuables à une maladie moins grave comme un rhume ou une bronchite. Cependant, ils pourraient être secondaires à une affection plus sérieuse. La conduite thérapeutique étant tout à fait différente dans les deux situations, il est essentiel de préciser le diagnostic.

Souvent, les renseignements obtenus au cours de l'anamnèse et de l'examen physique ne suffisent pas pour poser un diagnostic sûr permettant de choisir un traitement approprié. Pour cette raison, le clinicien complète souvent l'investigation par des analyses de laboratoire dont il doit savoir interpréter correctement les résultats. Ce chapitre vise à définir les caractéristiques d'un examen diagnostique et à introduire les principes servant à en interpréter les résultats.

7.1 La classification des résultats d'un examen diagnostique

Les examens diagnostiques peuvent être utilisés dans un contexte de diagnostic, de dépistage ou d'évaluation de l'efficacité thérapeutique. Selon le type d'examen effectué, les résultats sont qualitatifs ou quantitatifs et peuvent être exprimés sur une échelle nominale, ordinale ou par intervalle (*voir le tableau 2.2*).

EXEMPLE 7.1

- Échelle nominale – La culture du *Mycobacterium tuberculosis* se révèle positive ou négative. Une radiographie pulmonaire sera jugée normale ou anormale. La douleur est rétrosternale ou non.
- Échelle ordinale – Le pouls radial est lent, normal ou rapide. Le réflexe cutané plantaire est absent, normal ou vif. L'intensité des souffles cardiaques systoliques est décrite en utilisant une échelle graduée de 1 à 6.
- Échelle par intervalle – La pression artérielle diastolique est mesurée à 92 mm Hg, à 98 mm Hg ou à 110 mm Hg. La glycémie est mesurée à 3,9 mmol/l, à 5,8 mmol/l ou à 8,3 mmol/l. Ces observations cliniques sont des variables continues et s'expriment sur une échelle par intervalle.

On utilise habituellement le test diagnostique auprès de sujets symptomatiques, chez qui l'on cherche à préciser la cause des problèmes. Le test de dépistage, lui, recherche une maladie non décelable cliniquement chez des sujets asymptomatiques. Cette distinction se répercute sur plusieurs plans : en situation de dépistage, on cherchera un test particulièrement simple, peu coûteux et peu effractif.

7.1.1 La comparaison d'un examen diagnostique à un examen de référence

Pour établir un diagnostic, il serait souhaitable de toujours recourir à un **examen de référence**, soit la méthode diagnostique la plus valide qu'il soit possible d'utiliser dans une situation clinique donnée. Malheureusement, l'examen de référence peut présenter certains risques pour la santé, être techniquement difficile à effectuer, coûteux ou encore impossible à réaliser, d'où la nécessité de procéder à l'étude de tests diagnostiques alternatifs.

EXEMPLE 7.2

- L'examen anatomopathologique de la vésicule biliaire est l'examen de référence permettant de conclure à une cholécystite. Par contre, l'ouverture chirurgicale de la paroi abdominale de tous les patients présentant une douleur à l'hypocondre droit constitue un risque évident pour la santé. L'échographie abdominale constitue une option diagnostique moins effractive.
- La coronarographie permet de visualiser le degré d'obstruction des artères coronaires. Elle présente un risque de complications non négligeable pour le patient.

EXEMPLE 7.2 (suite)

C'est l'examen de référence qui permet d'évaluer la validité d'examens alternatifs comportant moins de risques, dans le diagnostic d'une maladie coronarienne.

- L'examen histopathologique du cerveau à l'autopsie est l'examen de référence permettant de confirmer le diagnostic de la maladie d'Alzheimer. On ne s'étonne donc pas que beaucoup d'études se penchent sur l'évaluation de moyens diagnostiques de rechange!

7.1.2 La classification des résultats

Pour évaluer la validité d'un nouvel **examen diagnostique**, il est nécessaire de comparer simultanément le résultat obtenu à cet examen au résultat d'un examen de référence à l'aide d'un tableau de contingence pour chacun des individus testés.

Pour illustrer cette procédure, nous utiliserons la situation la plus simple, celle où l'examen diagnostique et l'examen de référence sont appliqués à tous les sujets et nous donnent tous deux des résultats dichotomiques. Nous exprimerons les résultats de l'examen diagnostique ainsi: positif (E+) ou négatif (E-); ceux du test de référence sont indiqués ainsi: malades (M+) ou non malades (M-). Le tableau 7.1 illustre la façon standard de dresser un tableau de contingence. On remarque qu'il existe quatre combinaisons de résultats. Deux catégories pour lesquelles les résultats du nouvel examen sont en accord avec le résultat de l'examen de référence: le vrai positif (VP) en présence de la maladie et le vrai négatif (VN) en son absence. Les deux autres résultats du nouveau test diagnostique sont erronés: le faux positif (FP), test diagnostique positif alors que la maladie est réellement absente, et le faux négatif (FN), test diagnostique négatif alors que la maladie est réellement présente.

TABLEAU 7.1 ■ La classification des résultats d'un examen diagnostique

	M +	M -	Total
E +	VP	FP	VP + FP
E -	FN	VN	FN + VN
Total	VP + FN	FP + VN	

E: examen diagnostique M: examen de référence

En général, l'examen diagnostique est plus simple, moins coûteux et moins effrayant à pratiquer que l'examen de référence. C'est le cas, par exemple, du dosage des d-dimères comparativement à l'angioscanner thoracique pour le diagnostic de l'embolie pulmonaire. Mais Lisa se demande dans quelle mesure les renseignements qu'elle obtiendra avec le dosage des d-dimères lui permettront de poser un diagnostic juste.

Pour bien cerner les défis que pose l'utilisation judicieuse de tests diagnostiques, il importe de se référer au concept de validité des tests. La validité, ou l'absence de biais, renvoie au degré avec lequel un test mesure ce qu'il doit effectivement mesurer.

On distingue habituellement deux niveaux de validité: la validité interne ou intrinsèque (mesure exacte dans la population étudiée) et la validité externe ou extrinsèque (capacité de généraliser les résultats à d'autres populations).

7.2 La validité intrinsèque : sensibilité (SE) et spécificité (SP)

La sensibilité mesure la capacité d'un test à repérer adéquatement les sujets réellement malades alors que la spécificité mesure la capacité d'un test à reconnaître adéquatement les sujets qui, en réalité, ne sont pas malades. La sensibilité et la spécificité sont les deux paramètres qui caractérisent la **validité intrinsèque** d'un examen diagnostique. Comme tout clinicien, Lisa doit en apprendre davantage sur la valeur de ces paramètres pour interpréter correctement la signification d'un examen dont le résultat est négatif ou positif.

La sensibilité et la spécificité sont des paramètres qui doivent être estimés dans des études spécialement conçues à cette fin. Ces études, conceptuellement simples, s'avèrent souvent difficiles à réaliser. La procédure consiste à déterminer, en utilisant la méthode diagnostique la plus valide (soit l'examen de référence), lesquels des sujets d'un groupe à l'étude sont malades et lesquels sont sains. Puis, les deux groupes de sujets subissent l'examen diagnostique. Lorsque l'étude est bien menée (*voir à ce sujet le chapitre 8*), la sensibilité et la spécificité d'un examen diagnostique sont des valeurs qui demeurent généralement constantes. De plus en plus, lorsque l'examen de référence est très intrusif et comporte des risques jugés éthiquement non acceptables, les études de nouveaux tests diagnostiques utiliseront un devis différent. Par exemple, les participants présentant des résultats négatifs à l'issue du nouveau test pourront être suivis dans le temps pour vérifier que la maladie d'intérêt n'apparaît pas. Bien entendu, cette stratégie ne peut être envisagée que dans les cas de maladies qui sont presque invariablement symptomatiques et qui conduiront les participants à consulter et à l'établissement d'un diagnostic.

EXEMPLE 7.3

Une étude visant à évaluer différentes combinaisons de tests pouvant mener à un diagnostic d'embolie pulmonaire de manière efficiente a été conduite en Suisse¹. Les participants ont été recrutés de manière consécutive dans trois urgences d'hôpitaux universitaires. Ils étaient admissibles s'ils se présentaient pour une dyspnée ou une douleur thoracique d'apparition récente, sans autre cause évidente.

Le tableau de contingence suivant résume les résultats pour 629 participants ayant eu un dosage de d-dimères et pour qui il a été possible de poser ou d'éliminer un diagnostic d'embolie pulmonaire.

	M +	M -	Total
d-dimères +	117	292	409
d-dimères -	0	220	220
Total	117	512	629

Pour le tableau, les données des participants avec risque faible et moyen d'embolie pulmonaire selon l'échelle de Genève, avec suivi à trois mois sans anticoagulation pour autre raison, ont été incluses.

Source : Adapté de Perrier, A., Roy, P.M., Sanchez, O., Le Gal, G. et autres (2005). Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med*, 352(17), 1760-1768.

7.2.1 Le calcul de la sensibilité et de la spécificité

La **sensibilité** (SE) d'un test correspond à la probabilité que celui-ci produise un résultat positif chez le sujet malade. Elle nous permet d'évaluer la capacité d'un test à reconnaître adéquatement les sujets réellement malades. Elle est déterminée par la proportion de personnes malades ayant obtenu un résultat d'examen positif; cette proportion s'exprime de la façon suivante :

$$SE = \frac{VP}{VP + FN}$$

où VP représente les vrais positifs, et FN, les faux négatifs.

En présence de la maladie, un examen très sensible est presque toujours positif.

EXEMPLE 7.4

Dans l'exemple 7.3, la sensibilité du dosage des d-dimères est de 100 % (SE = 117/117). Cela signifie que les 117 participants chez qui un diagnostic d'embolie pulmonaire a été confirmé ont obtenu un dosage de d-dimères se situant au-dessus de la valeur seuil pour un test positif. La proportion de faux négatifs, soit la proportion de participants identifiés par le dosage de d-dimères comme n'ayant pas la maladie alors qu'ils en sont atteints en réalité, s'établit à 0 % (0/117).

Plus simplement, la proportion de faux négatifs est obtenue par :

$$1 - SE$$

La **spécificité** (SP) d'un test correspond à la probabilité que celui-ci produise un résultat négatif chez le sujet non malade. Elle permet d'évaluer la capacité d'un test à reconnaître adéquatement les sujets qui ne sont pas malades. Elle est déterminée par la proportion de personnes non malades ayant obtenu un résultat d'examen négatif; cette proportion s'exprime de la façon suivante :

$$SP = \frac{VN}{VN + FP}$$

où VN représente les vrais négatifs, et FP, les faux positifs.

EXEMPLE 7.5

Dans l'exemple 7.3, la spécificité du dosage des d-dimères de 43 % (SP = 220/512) signifie que, parmi les 512 individus n'ayant pas de maladie coronarienne, 220 ont obtenu un résultat de dosage de d-dimères se situant en dessous de la valeur seuil. La proportion de faux positifs, soit la proportion de patients identifiés par le dosage de d-dimères comme ayant possiblement la maladie alors qu'ils n'en sont pas atteints, s'établit à 57 % (292/512). Plus simplement, la proportion de faux positifs est obtenue par :

$$1 - SP$$

Ainsi, pour désigner l'examen diagnostique le plus avantageux dans une situation clinique donnée, Lisa doit d'abord prendre en considération les caractéristiques intrinsèques de l'examen, soit sa sensibilité et sa spécificité.

7.2.2 Le choix d'un examen sensible ou spécifique

Le choix d'un examen sensible

On tend à choisir un examen présentant une excellente sensibilité au début de la démarche diagnostique, puisqu'il s'agit alors d'exclure une ou plusieurs maladies de la liste des possibilités envisagées. Par exemple, pour une maladie donnée, l'obtention d'un résultat négatif à un examen sensible à 100 % permettra au clinicien d'exclure en toute certitude cette possibilité diagnostique (aucun patient réellement malade n'obtiendra un test négatif). Il existe toutefois très peu d'examens sensibles à 100 %. Par exemple, d'autres études évaluant la performance du dosage des d-dimères dans le diagnostic de l'embolie pulmonaire ont trouvé une sensibilité inférieure à 100 % (mais tout de même très élevée, soit de plus de 95 %, pour les méthodes ultrasensibles)².

On utilise un examen très sensible lorsque les conséquences d'un résultat faussement négatif sont importantes ou inacceptables. C'est le cas, par exemple, lorsque la maladie recherchée est grave et qu'il existe un traitement efficace et accessible.

Le choix d'un examen spécifique

L'examen diagnostique présentant une excellente spécificité sera choisi pour confirmer une maladie dont on soupçonne fortement la présence à la suite d'une investigation antérieure. Un examen spécifique à 100 % ne conduit à aucun résultat faussement positif, ce qui permet de confirmer un diagnostic avec certitude. Évidemment, très peu d'examens sont spécifiques à 100 %.

Un examen très spécifique s'avère par ailleurs essentiel lorsque les conséquences potentielles d'un résultat faussement positif sont inacceptables. Par exemple, lorsque le traitement comporte des effets secondaires majeurs, physiques ou psychologiques (par exemple, la radiothérapie et la chimiothérapie utilisées dans le traitement du cancer), ou lorsque le diagnostic lui-même est celui d'une maladie mortelle.

7.2.3 L'influence du seuil de positivité

On a vu qu'il était plus simple d'évaluer la validité intrinsèque d'un examen diagnostique si son résultat est dichotomique. En conséquence, on voudra fréquemment convertir le résultat d'un test s'exprimant sur une échelle ordinale ou par intervalle en résultat dichotomique; il s'agira de fixer un seuil à partir duquel le résultat sera considéré comme anormal (positif). C'est le **seuil de positivité** (aussi appelé seuil de normalité) d'un examen.

Il existe plusieurs façons de déterminer un seuil de positivité. Il est notamment possible d'utiliser la distribution de fréquences des valeurs du résultat d'un examen diagnostique dans une population d'individus non malades. Les valeurs extrêmes de cette distribution (par exemple, les valeurs situées au-dessus du 95^e percentile) peuvent être considérées comme inhabituelles ou anormales. Évidemment, si une

telle façon de faire correspond à une définition statistique du seuil de positivité (de normalité), elle ne repose néanmoins sur aucune base biologique. La fixation d'un tel seuil implique également que, pour tout examen diagnostique, 5 % des individus seraient systématiquement classifiés comme « anormaux ».

EXEMPLE 7.6

On peut considérer une croissance standard *in utero* comme un signe de développement normal du fœtus. Cette croissance peut être évaluée, notamment au cours du troisième trimestre de la grossesse, par des mesures échographiques permettant d'estimer le poids du fœtus. Les différentes valeurs du poids s'expriment sur une échelle par intervalle. Si l'on considère que la croissance d'un fœtus est normale lorsque son poids estimé se situe entre le 10^e et le 90^e percentile de la distribution de fréquences du poids — la mesure du poids étant ici l'examen diagnostique —, l'examen est considéré comme positif (E+) lorsque le poids estimé se situe au-dessus du 90^e percentile ou au-dessous du 10^e percentile; il est négatif (E-) lorsque le poids estimé se situe entre le 10^e percentile et le 90^e percentile. Il s'agit ici d'une définition statistique du seuil de positivité; elle peut aider à reconnaître les fœtus nécessitant une surveillance plus étroite. Toutefois, l'évolution des percentiles dans le temps (par exemple, un fœtus évalué au 30^e percentile qui serait au 5^e percentile un mois plus tard) sera plus utile qu'une mesure prise isolément. Par ailleurs, le fait que le poids de certains nouveau-nés, en parfaite santé, se situe aux extrêmes des courbes de poids illustre bien les limites d'une définition statistique de la normalité.

Le seuil de positivité correspond parfois à une certaine valeur au-dessus de laquelle le risque de contracter une maladie donnée est considéré comme excessif. Ce seuil peut également correspondre à la valeur à partir de laquelle l'instauration d'un traitement est associée à des gains de santé (par exemple, une diminution de la mortalité ou de l'incidence). Par conséquent, le seuil de positivité doit être modifié à mesure qu'évo-luent les traitements offerts et que le clinicien en est informé.

EXEMPLE 7.7

Les résultats de mammographie sont habituellement rapportés selon le système BI-RADS (*Breast Imaging Reporting and Data System*), sur une échelle de 1 à 5, alors qu'un résultat égal à 0 indique une évaluation incomplète. Des scores de 1 et 2 sont considérés comme normaux ou associés à une pathologie bénigne. Un score de 3 est considéré comme équivoque et requiert un contrôle rapproché. Un score de 4 ou 5 indique un risque élevé de cancer, et une biopsie des lésions détectées s'avère indiquée.

EXEMPLE 7.8

La pression artérielle diastolique (PAD) est une variable continue dont les différentes valeurs s'expriment sur une échelle par intervalle. Le risque de décès par maladie cardiovasculaire augmente de manière log-linéaire avec la valeur de la PAD, sans valeur seuil. La majorité des experts considèrent que les bénéfices de traitement excèdent les risques lorsque la PAD dépasse 90-100 mm Hg; pour

EXEMPLE 7.8 (suite)

cette raison, un traitement est habituellement recommandé³. La mesure de la PAD devient ici l'examen diagnostique. L'examen est considéré comme positif (E+) lorsque la PAD est supérieure à 90 mm Hg.

Pour déterminer un seuil de positivité d'une manière plus avantageuse, il faudra fixer celui-ci en fonction de l'utilité clinique recherchée. En effet, lorsque les valeurs du résultat d'un examen sont exprimées sur une échelle par intervalle, la sensibilité et la spécificité de l'examen varient selon le seuil de positivité retenu. Une bonne validité intrinsèque découle d'une bonne sensibilité et d'une bonne spécificité. Toutefois, comme il est généralement impossible d'obtenir à la fois une très forte sensibilité et une très forte spécificité — une augmentation de l'une entraîne une baisse de l'autre —, un compromis s'impose. Le choix d'un seuil de positivité peut donc être arrêté en fonction de l'objectif clinique: si l'on cherche avant tout à exclure un diagnostic, on préférera un seuil associé à une bonne sensibilité; par contre, si l'on cherche essentiellement à confirmer un diagnostic, on optera pour un seuil présentant une bonne spécificité.

Le choix du seuil de positivité: l'exemple du dépistage du cancer du col de l'utérus

Le cancer du col de l'utérus demeure un problème de santé publique important, étant le quatrième cancer le plus fréquent chez les femmes dans le monde, mais le deuxième plus courant dans les pays moins développés⁴. La cytologie cervicale a contribué à diminuer l'incidence et la mortalité associées au cancer du col utérin dans les pays plus développés qui ont pu mettre en place l'infrastructure et l'assurance de qualité nécessaire à la bonne performance de ce test, ce qui n'a pas été possible dans les pays moins développés.

On sait maintenant que les virus du papillome humain oncogènes, ou VPH, sont la cause nécessaire du cancer du col utérin. Des études ont aussi démontré que des tests de détection d'ADN de VPH dans les sécrétions cervicales pouvaient améliorer le dépistage du cancer du col. Certains tests de détection de VPH ont été mis au point dans le but précis d'être utilisables dans les conditions qui prévalent dans les pays moins développés, comme le test careHPV⁵.

Kang et ses collègues⁶ ont mené une étude afin de déterminer le seuil de positivité optimal pour ce test. L'étude a été réalisée à partir d'un échantillon de 9 785 femmes ayant participé à 2 recherches conduites entre 2007 et 2011 en Chine rurale. Chacune des participantes a fourni des échantillons vaginaux qui ont été analysés avec différents tests VPH. Elles ont ensuite eu une inspection du col après application d'acide acétique, ce qui permet de visualiser les précancers et les cancers et de diriger les biopsies. L'analyse des biopsies du col a été l'examen de référence permettant de déterminer la présence ou l'absence de la maladie.

Le tableau 7.2 résume les résultats obtenus en analysant les échantillons vaginaux recueillis par un professionnel de la santé avec le test careHPV. Le but du test est de détecter les cancers du col ou les précurseurs de cancers, appelés CIN3 (*cervical intra-epithelial neoplasia grade 3*).

TABLEAU 7.2 ■ Les résultats du test careHPV

Seuil de positivité RLU	Nombre de femmes positives pour le VPH (%)	Sensibilité (%)	Spécificité (%)
0,50	1982 (20,3)	98,85	80,75
1,00	1401 (14,3)	95,28	86,75
1,74	1206 (12,3)	94,49	88,76
2,00	1165 (11,9)	93,70	89,17
4,00	999 (10,2)	88,98	90,83
10,00	820 (8,4)	83,46	92,61

Source: Adapté de Kang, L.N., Jeronimo, J., Qiao, Y.L, et autres (2014). Optimal positive cutoff points for careHPV testing of clinician – and self-collected specimens in primary cervical cancer screening: An analysis from rural China. *J Clin Microbiol*, 52(6), 1954-. RLU : relative light units.

À titre d'exemple, si le seuil de positivité est fixé à 1,74 RLU, 94,49 % des participantes atteintes d'un cancer du col ou d'un CIN3 auront un test VPH positif, ce qui correspond à la sensibilité (SE) de l'examen pour ce seuil. On observe également que 88,76 % des participantes non atteintes de cancer ou de CIN3 obtiendront un test VPH négatif, ce qui correspond à la spécificité (SP) de l'examen pour ce seuil. Or, dans un contexte où l'on veut dépister le plus grand nombre de participantes atteintes d'un cancer du col, on choisira un examen qui présente une plus grande sensibilité, en s'assurant cependant de conserver une spécificité adéquate. Si le seuil de positivité est fixé à 10,00 RLU, la sensibilité diminue à 83,46 %, et la spécificité augmente à 92,61 %. Ainsi, avec un seuil de 1,74 RLU, la proportion de faux positifs est de 21,24 % ($1 - SP$) comparativement à 7,4 % avec un seuil à 10,00 RLU. Une proportion plus élevée d'individus est considérée à tort comme étant malade, vit inutilement de l'anxiété et est soumise à des examens de confirmation coûteux et effrayants. Par contre, avec un seuil de 1,74 RLU, la proportion de faux négatifs est de 5,51 % ($1 - SE$) comparativement à 16,54 % avec un seuil à 10,00 RLU. Avec une sensibilité plus faible, la proportion de faux négatifs augmente, et une proportion plus grande d'individus reçoit un résultat normal, alors que ces personnes sont en réalité atteintes de la maladie.

Le choix d'un seuil de positivité impose donc un compromis et doit reposer sur l'objectif d'utilisation clinique de l'examen. Dans le présent contexte, comme il s'agit de dépister le plus grand nombre de femmes risquant d'être atteintes d'un cancer du col de l'utérus, on recherche un examen ayant une bonne sensibilité. De même, afin de réduire le nombre d'examens faussement positifs, l'examen doit également présenter une bonne spécificité.

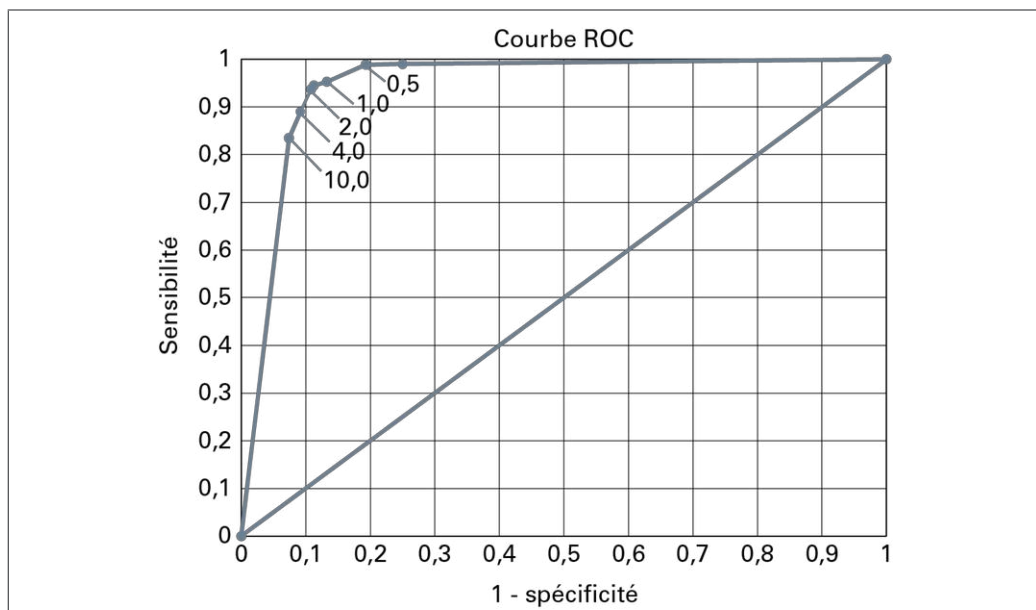
En résumé, la sensibilité et la spécificité d'un test, en tant que caractéristiques stables, ne se modifient que si l'on déplace le seuil de positivité. Ainsi, lorsqu'on augmente la sensibilité, on diminue nécessairement la spécificité; à l'inverse, quand on augmente la spécificité, on diminue naturellement la sensibilité.

La courbe récepteur opérateur (ROC)

Une manière d'illustrer la façon dont la sensibilité et la spécificité varient selon la valeur seuil est d'utiliser une courbe récepteur opérateur, mieux connue sous le nom de courbe ROC de l'acronyme anglais signifiant *receiver-operator curve* (nommée ainsi parce qu'elle a d'abord servi à distinguer les signaux de radar du bruit de fond sur les navires et dans les sous-marins). Pour construire ce type de courbe, on doit utiliser un graphique avec, en ordonnée, la proportion de vrais positifs (sensibilité) et, en abscisse, la proportion de faux positifs ($1 - \text{spécificité}$). On construit la courbe en reportant ces valeurs sur le graphique pour différents seuils. La courbe pour un test sans utilité s'approchera d'une ligne droite diagonale. La courbe ROC pour un test qui aidera à bien différencier les malades des non-malades aura une forme convexe, le sommet de la courbe s'approchant du coin supérieur gauche du graphique. Cette courbe permet donc de comparer des tests : le test doté de la meilleure discrimination affichera la plus grande aire sous la courbe. Celle-ci permet aussi de sélectionner la meilleure valeur seuil, qui sera celle située le plus près du coin supérieur gauche du graphique.

La figure 7.1 montre la courbe ROC issue des données de l'étude de Kang et de ses collaborateurs sur le point de coupure optimal du test careHPV. On voit la forme très convexe de la courbe confirmant que le test permet de bien distinguer les malades des non-malades. Le point de la courbe situé le plus près du coin supérieur gauche, soit 2,0, est celui qui donnera le meilleur compromis sur le plan de la sensibilité et de la spécificité.

FIGURE 7.1 ■ La courbe ROC



Source: Adapté de Kang, L.N. et autres (2014). Optimal positive cutoff points for careHPV testing of clinician – and self-collected specimens in primary cervical cancer screening: An analysis from rural China. *J Clin Microbiol*, 52(6),1954-.

7.3 La validité extrinsèque: les valeurs prédictives positives et négatives

On établit la sensibilité et la spécificité d'un examen à partir de la connaissance du diagnostic juste. Or, en pratique, le clinicien ne connaît pas ce diagnostic final et confirmé lorsqu'il prescrit le test. C'est à partir de la collecte d'informations par questionnaire et examen physique que le clinicien estime la probabilité que ce patient soit sain ou malade. Cette probabilité, établie avant toute demande d'examen diagnostique, correspond à la prévalence de la maladie dans une population donnée. Le clinicien appuie donc son évaluation du risque que présente son patient tant sur les données tirées de la littérature que sur son expérience clinique. Cette probabilité pourra par exemple être faible, modérée ou forte, soit de 5 %, de 30 % ou de 90 %.

De nombreuses études ont tenté d'estimer la probabilité d'embolie pulmonaire chez des patients manifestant des symptômes pouvant être compatibles avec cette maladie. Des règles de prédiction cliniques ont été développées, puis validées. Celles-ci permettent de déterminer le risque d'embolie pulmonaire selon une série de critères facilement obtenus à l'histoire clinique et durant un bref examen. Par exemple, le score de Genève⁷ simplifié attribue un point pour chacune des situations suivantes :

- âge de plus de 65 ans ;
- antécédent de thrombophlébite ou d'embolie pulmonaire ;
- antécédent de chirurgie ou de fracture récente ;
- condition maligne active ;
- douleur unilatérale à un membre inférieur ;
- hémoptysie ;
- tachycardie ;
- douleur à la palpation d'un membre inférieur ou œdème unilatéral.

La prévalence d'embolie pulmonaire varie selon le nombre de critères, comme le montre le tableau 7.3.

TABLEAU 7.3 ■ La prévalence de l'embolie pulmonaire selon le nombre de critères du score simplifié de Genève

Nombre de critère(s) présent(s)	Prévalence d'embolie pulmonaire
0-1	8%
2-4	29%
5 et plus	64%

Source: Adapté de Klok, F.A., Mos, I.C., Nijkeuter, M. et autres (2008). Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism. *Arch Intern Med*, 168(19), 2131-2136.

EXEMPLE 7.9

Dans la mise en situation, on retient que Paule a un pouls rapide et a subi une chirurgie récemment. Elle présente donc deux critères du score de Genève simplifié présenté dans le tableau 7.3. Ainsi, avant même de prescrire un examen, Lisa peut estimer à 29 % la probabilité que Paule soit atteinte d'une embolie pulmonaire.

En clinique, on recourt aux examens diagnostiques pour obtenir des renseignements supplémentaires permettant de réviser l'estimation de la probabilité qu'un patient soit atteint ou non de la maladie. On obtient ce gain d'information à partir du calcul des valeurs prédictives.

7.3.1 Les valeurs prédictives positives et négatives

Exprimées en pourcentage, les valeurs prédictives — calculées à partir du tableau de contingence — représentent la probabilité qu'un individu soit malade ou non selon le résultat de l'examen. Une **valeur prédictive positive (VPP)** indique la probabilité que l'individu soit réellement malade lorsque le résultat d'un examen s'avère positif. En d'autres mots, il s'agit de la proportion des individus réellement atteints de la maladie, soit les vrais positifs (VP), parmi l'ensemble des individus dont le résultat est positif, soit les vrais positifs (VP) et les faux positifs (FP).

$$VPP = \frac{VP}{VP + FP}$$

La probabilité d'observer un résultat faussement positif ($1 - VPP$) correspond à la proportion des résultats faussement positifs obtenus parmi l'ensemble des individus dont le résultat est positif. Ainsi, même si les individus de ce groupe ne sont pas malades, le résultat de ce groupe sera positif.

En revanche, en présence d'un examen dont le résultat est négatif, le clinicien doit estimer la probabilité que le patient soit réellement sain. Cette probabilité, appelée **valeur prédictive négative (VPN)**, est déterminée par la proportion des individus qui ne sont pas atteints de la maladie, soit les vrais négatifs (VN), parmi l'ensemble des individus dont le résultat est négatif, soit les vrais négatifs (VN) et les faux négatifs (FN).

$$VPN = \frac{VN}{VN + FN}$$

La probabilité d'observer un résultat faussement négatif ($1 - VPN$) correspond à la proportion des résultats faussement négatifs obtenus parmi l'ensemble des individus dont le résultat est négatif. Forcément, une certaine proportion d'individus malades obtiendra un résultat négatif.

7.3.2 Le calcul des valeurs prédictives

Contrairement à la sensibilité et la spécificité, la prévalence de la maladie influe sur la VPP et la VPN d'un test. Pour une sensibilité et une spécificité données, la VPP diminue, et la VPN augmente à mesure que la prévalence décroît.

Il est possible de calculer la VPP et la VPN si l'on connaît la prévalence de la maladie, la sensibilité et la spécificité de l'examen diagnostique. Il s'agit simplement de recréer un tableau de contingence à partir des informations connues.

Considérant que la prévalence (ou probabilité prétest) de la maladie de Paule est estimée à 29 %, que les tests les plus sensibles pour les d-dimères ont, en moyenne, une sensibilité de 97 % et une spécificité de 40 %⁸, Lisa peut effectuer le calcul de la VPP et de la VPN en six étapes.

1. Elle choisit de façon arbitraire une taille d'échantillon, par exemple 1 000 individus.
2. À partir de la prévalence, elle calcule le nombre d'individus malades et non malades, soit :
malades : $(1\ 000)(0,29) = 290$;
non malades : $1\ 000 - 290 = 710$.
3. À partir de la sensibilité de l'examen, elle calcule, parmi les individus malades, le nombre d'individus dont le résultat est positif et le nombre d'individus dont le résultat est négatif, soit :
d-dimères + : $(290)(0,97) = 281$; d-dimères - : $290 - 281 = 9$.
4. À partir de la spécificité de l'examen, elle calcule aussi, parmi les individus non malades, le nombre d'individus dont le résultat est positif et le nombre d'individus dont le résultat est négatif, soit :
d-dimères - : $(710)(0,40) = 284$; d-dimères + : $710 - 284 = 426$.
5. Ensuite, elle calcule le nombre total d'individus dont le résultat est positif et le nombre total d'individus dont le résultat est négatif, soit :
d-dimères + : $281 + 426 = 707$; d-dimères - : $9 + 284 = 293$.
6. Enfin, elle calcule la probabilité d'observer un résultat vraiment positif parmi l'ensemble des résultats positifs (VPP) et la probabilité d'observer un résultat vraiment négatif parmi l'ensemble des résultats négatifs (VPN), soit :
VPP : $281/707 = 0,40$; VPN : $284/293 = 0,97$.

Le tableau 7.4 résume ces différentes étapes.

TABLEAU 7.4 ■ La classification des résultats du dosage des d-dimères (prévalence de 29 %)

	M +	M -	Total
d-dimères +	281	426	707
d-dimères -	9	284	293
Total	290	710	1 000

Comme la VPP est de 40 %, c'est donc dire que, si le dosage des d-dimères se situe au-dessus du seuil de positivité, le gain d'information est tout de même intéressant. En effet, la probabilité que Paule soit atteinte d'une embolie pulmonaire — de 29 % avant l'examen (prévalence) — augmente à 40 % à la suite du dosage des d-dimères. Étant donné la gravité du diagnostic envisagé, il est recommandé de procéder à des examens additionnels dans cette situation.

Par ailleurs, comme la VPN est de 97 %, si le résultat de l'examen se révèle négatif, la probabilité que Paule ne souffre pas d'une embolie pulmonaire est de 97 %. Dans ce contexte, la VPN ajoute beaucoup d'information à ce qui est déjà connu, à savoir qu'avant l'examen, la probabilité que Paule ne soit pas atteinte d'une embolie pulmonaire était de 71 % (100 % – 29 %) alors qu'après un dosage négatif, elle grimpe à 97 %. Dans ce contexte, on pourrait rassurer Paule en lui disant qu'elle n'a pas d'embolie pulmonaire, et il ne serait pas nécessaire de demander des examens plus effractifs.

7.3.3 L'influence de la prévalence et de la validité intrinsèque

La valeur prédictive est le résultat de l'interaction entre la prévalence de la maladie, la sensibilité de l'examen diagnostique et sa spécificité.

L'influence de la prévalence

En matière d'examen diagnostique dont la validité intrinsèque est connue (sensibilité et spécificité), l'interprétation d'un résultat positif ou négatif doit tenir compte du contexte clinique. En effet, la prévalence de la maladie (probabilité de la maladie avant l'examen) varie grandement selon les caractéristiques des patients.

EXEMPLE 7.10

Lisa se souvient de Henri, un professeur âgé de 70 ans, qui présentait un essoufflement et des hémoptysies. Il se plaignait d'une douleur au mollet gauche, et l'examen avait mis en évidence une tachycardie, une douleur à la palpation et un œdème du membre inférieur gauche. Henri présentait cinq critères du score de Genève modifié. La probabilité qu'il souffre d'une embolie pulmonaire était donc de 64 % (voir le tableau 7.3). Dans ce contexte, le dosage des d-dimères apporterait peu de renseignements supplémentaires, comme l'indique le tableau de contingence 7.5.

TABLEAU 7.5 ■ La classification des résultats du dosage des d-dimères (prévalence de 64 %)

	M +	M –	Total
d-dimères +	621	216	837
d-dimères –	19	144	163
Total	640	360	1 000

EXEMPLE 7.10 (suite)

La prévalence de la maladie étant de 64 %, la VPP atteindrait 74 % (621/837) si le dosage des d-dimères se situait au-dessus du seuil de positivité. Cela ajoute peu d'information étant donné que la prévalence en elle-même constituait une indication de procéder à des examens plus poussés. Si le résultat avait été négatif, la VPN aurait été de 88 % (144/163), indiquant ainsi que la probabilité qu'Henri soit réellement atteint d'une embolie aurait atteint 12 % (100 % – 88 %). Ainsi, même en présence d'un résultat négatif, cette probabilité ne rassure pas complètement Lisa, et elle demande donc des tests supplémentaires.

EXEMPLE 7.11

Lisa se souvient également de Constance, jeune policière de 30 ans, qui se plaignait d'une douleur thoracique de faible intensité du côté droit; survenue en l'absence de traumatisme et d'une durée très variable, cette douleur s'atténuait après quelques exercices d'étirement. À l'examen, Lisa avait noté une tachycardie, sans aucune autre anomalie. Constance ne présentait qu'un des critères du score de Genève modifié et avait donc une probabilité d'embolie pulmonaire de 8 % (voir le tableau 7.3). En tenant compte de cette probabilité, le résultat du dosage des d-dimères ne lui serait pas d'une très grande aide.

Comme l'indique le tableau de contingence 7.6, le calcul des valeurs prédictives lui donne raison.

TABLEAU 7.6 ■ La classification des résultats du dosage des d-dimères (prévalence de 8 %)

	M +	M –	Total
d-dimères +	78	552	630
d-dimères –	2	368	370
Total	80	920	1 000

La prévalence de la maladie étant de 8 %, la VPP aurait été de 12 % (78/630), si le résultat du dosage des d-dimères avait été au-dessus du seuil de positivité. Malgré un résultat positif, ce faible gain d'information n'aurait pas beaucoup aidé Lisa à planifier les prochaines étapes de l'investigation. Dans l'éventualité d'un résultat négatif, la VPN aurait été de 99 % (368/370), indiquant ainsi une probabilité de 1 % (100 % – 99 %) d'embolie pulmonaire. Évidemment, cette très faible probabilité aurait conforté Lisa dans sa décision de ne pas proposer de tests supplémentaires ou de traitement à Constance pour ce problème de santé.

Ces deux exemples illustrent clairement qu'en règle générale, lorsque la prévalence de la maladie est très forte ou très faible, le gain d'information que procurent les examens diagnostiques s'avère peu utile à la décision d'offrir ou non un traitement.

Cependant, les examens diagnostiques offrent un gain d'information intéressant dans un contexte où la prévalence estimée de la maladie ne suffit pas à orienter la conduite, comme c'est le cas pour Paule.

L'influence de la sensibilité et de la spécificité sur les valeurs prédictives

La spécificité de l'examen — contrairement à sa sensibilité — influe fortement sur la valeur prédictive positive. Pour une prévalence donnée, plus la spécificité est élevée, plus la VPP est élevée. Il est facile de s'en convaincre. En effet, si l'on modifie la spécificité tout en assurant une sensibilité constante à 70 % et une prévalence également constante à 50 %, on observe que la spécificité et la VPP augmentent de façon concomitante; par ailleurs, la VPN est très peu modifiée. Les données présentées ci-dessous le démontrent.

SP (%)	80	85	90	95	99
VPP	78	82	88	93	99
VPN	73	74	75	76	77

Notons que, lorsque la prévalence de la maladie s'approche de 0, la VPP se rapproche elle aussi de 0, même lorsque la spécificité de l'examen est élevée. Cette situation prévaut dans un contexte de dépistage, où la prévalence de la maladie dans la population générale est souvent inférieure à 1 %.

EXEMPLE 7.12

Des intervenants en santé communautaire décident de mettre sur pied un programme de dépistage du cancer du sein. Pour ce faire, ils conviennent que les femmes qui obtiendront un résultat positif à la mammographie devront subir une biopsie afin de confirmer ou d'infirmer le diagnostic de cancer.

Dans le milieu où sera implanté le programme, le cancer du sein a une prévalence de 0,5 %. La sensibilité et la spécificité de la mammographie sont respectivement de 90 % et de 85 %⁹. Le tableau 7.7 présente les données permettant de calculer la VPP.

TABLEAU 7.7 ■ La classification des résultats de la mammographie

	M +	M -	Total
Mammographie +	4	149	153
Mammographie -	1	846	847
Total	5	995	1 000

On constate que, malgré l'excellente spécificité de la mammographie, la VPP n'est que de 2,6 % (4/153), ce qui signifie que 97,4 % des femmes dont le résultat sera positif devront inutilement subir une biopsie et vivre l'anxiété liée à l'annonce de

EXEMPLE 7.12 (suite)

la présence éventuelle d'un cancer. Pour éviter qu'une telle situation se produise, l'équipe décide de n'offrir le dépistage qu'aux femmes qui présentent une plus forte prévalence de la maladie (sous-groupes à risque), soit les femmes de 50 ans et plus, celles ayant des antécédents familiaux de cancer du sein, etc. Cette stratégie a pour effet d'améliorer la VPP de l'examen.

Inversement, c'est la sensibilité de l'examen qui influe surtout sur la valeur prédictive négative. Plus la sensibilité d'un examen est élevée, plus la valeur prédictive négative est élevée. Il est ici aussi facile de s'en convaincre. En effet, si l'on modifie la sensibilité tout en assurant une spécificité constante à 70% et une prévalence également constante à 50%, on observe que la sensibilité et la VPN augmentent de façon concomitante; quant à la VPP, la sensibilité influe très peu sur elle. Les données présentées ci-après le démontrent.

SE (%)	80	85	90	95	99
VPP	73	74	75	76	77
VPN	78	82	88	93	99

Notons que, même si l'examen est très sensible, un résultat négatif est souvent faussement négatif lorsque la prévalence de la maladie s'avère très élevée. Or, comme on l'a vu précédemment, dans un tel contexte, le résultat négatif d'un examen diagnostique incite rarement le clinicien à modifier sa conduite. Voilà donc un argument de plus pour renoncer à faire subir un examen au patient.

7.4 Les combinaisons d'examens

Dans certains contextes, le fait d'effectuer des examens en série ou en parallèle améliore la prise de décision. Dans une situation clinique non urgente, on procède le plus souvent à des examens en série; le résultat du premier examen devient le facteur déterminant qui décidera de la nécessité d'effectuer un examen complémentaire. Cette stratégie d'investigation tend à maximiser à la fois la VPP — les résultats positifs étant le plus souvent de vrais positifs — et la VPN.

7.4.1 La combinaison d'examens en parallèle

M. Bibeau, un quinquagénaire, se présente à l'urgence et se plaint de douleurs rétrosternales irradiantes au bras gauche; il transpire abondamment. Même si l'on administre d'emblée un traitement pour un infarctus, on doit également confirmer le diagnostic. Un électrocardiogramme (ECG) et un dosage des troponines sont demandés en priorité.

Que l'un ou l'autre des résultats s'avère positif confirmera le diagnostic, à plus forte raison si les deux le sont. En revanche, s'ils sont tous les deux négatifs, un autre

diagnostic devra être envisagé. Il s'agira de procéder à des **examens en parallèle**. On les pratique habituellement pour préciser rapidement un diagnostic, lorsqu'il est important qu'aucune possibilité ne soit négligée; on cherche donc à obtenir le moins de faux négatifs possible.

Imaginons qu'on reçoive à l'urgence 100 patients comme M. Bibeau et que, parmi eux, la prévalence de l'infarctus est de 75 %, la sensibilité de l'ECG de 90 % et sa spécificité de 75 %, la sensibilité du dosage des troponines I très sensibles est de 91 % et sa spécificité de 90,9 %¹⁰. Voici les tableaux 7.8 et 7.9 correspondants.

TABLEAU 7.8 ■ La classification des résultats de l'ECG

	M +	M -	Total
ECG +	68	6	74
ECG -	7	19	26
Total	75	25	100

TABLEAU 7.9 ■ La classification des résultats du dosage des troponines I très sensibles

	M +	M -	Total
Troponines I +	68	2	70
Troponines I -	7	23	30
Total	75	25	100

Avec l'ECG, on obtiendrait une VPN de 0,73 (19/26); avec le dosage très sensible de la troponine I, elle serait de 0,77 (23/30). Comment la sensibilité et la VPN seraient-elles augmentées si l'on effectuait les examens en parallèle?

On considère les examens menés en parallèle comme indépendants les uns des autres. Donc, pour deux examens, il y a quatre catégories possibles. Voici les étapes pour remplir le nouveau tableau.

1. On détermine le nombre de vrais positifs, soit ceux ayant obtenu un résultat positif à un des deux tests. Les 68 patients malades désignés comme positifs à l'ECG le resteraient. S'y ajouteraient 91 % des 7 faux négatifs de l'ECG que l'on aurait dépistés par le dosage très sensible de la troponine I, puisque la sensibilité de la troponine I est de 91 %. On ajoute donc 6 personnes ($7 \times 0,9$) dans la case des vrais positifs ($68 + 6 = 74$).
2. On détermine le nombre de faux positifs. Les six patients sains désignés comme positifs à l'ECG le resteraient. À ceux-ci s'ajouteraient 10 % des 19 vrais négatifs désignés comme positifs lors du dosage de la troponine I, puisque la spécificité de la troponine I est de 90 %. On ajoute donc 2 personnes ($19 \times 0,1$) dans la case des faux positifs ($6 + 2 = 8$).
3. Comme nous connaissons la prévalence de la maladie (75 %), nous pouvons maintenant calculer le nombre de faux négatifs ($75 - 74 = 1$) et le nombre de vrais négatifs ($25 - 8 = 17$). On se trouverait alors avec un tableau de contingence semblable au tableau 7.10.

TABLEAU 7.10 ■ La classification des résultats en parallèle de l'ECG et du dosage des troponines I

	M +	M -	Total
ECG + ou troponines I +	74	8	82
ECG - et troponines I -	1	17	18
Total	75	25	100

La sensibilité serait maintenant de 0,99 (74/75), et la VPN serait de 0,94 (17/18). On remarque que la spécificité, diminuée à 68 %, est le compromis nécessaire à l'obtention d'une sensibilité de 99 %.

7.4.2 La combinaison d'examens en série

Imaginons sensiblement la même situation : M. Bibeau se présente à la polyclinique et manifeste des symptômes moins typiques. On pourrait amorcer le bilan par un ECG et orienter la suite selon les résultats obtenus. Cette succession d'examens — le choix du deuxième examen dépend du résultat du premier — s'appelle **examens en série**. En ce qui concerne M. Bibeau, on ne poserait un diagnostic d'infarctus que si les deux examens étaient positifs. On obtiendrait donc, après l'ECG pratiqué au bureau, un tableau de contingence semblable au tableau 7.11.

TABLEAU 7.11 ■ La classification des résultats de l'ECG

	M +	M -	Total
ECG +	68	6	74
ECG -	7	19	26
Total	75	25	100

Cependant, étant donné que seules les personnes positives au premier examen seraient invitées à subir le second, le dosage de la troponine I serait effectué auprès d'une population présentant une prévalence d'infarctus bien plus importante. Cette prévalence serait probablement proche de la VPP de l'ECG, soit 0,92 (68/74); on obtiendrait ainsi un tableau de contingence similaire au tableau 7.12.

TABLEAU 7.12 ■ La classification des résultats du dosage des troponine I

	M +	M -	Total
Troponine I +	62	1	63
Troponine I -	6	5	11
Total	68	6	74

Les différents résultats aux deux examens seraient ceux que l'on pourrait regrouper dans un tableau de contingence semblable au tableau 7.13. Pour l'examen en série, on obtiendrait donc un autre tableau de contingence, semblable au tableau 7.14.

TABLEAU 7.13 ■ La classification des résultats de l'ECG et du dosage des troponines I

	M +	M –	Total
ECG + et troponine I +	62	1	60
ECG + et troponine I –	6	5	14
ECG –	7	19	26
Total	75	25	100

TABLEAU 7.14 ■ La classification des résultats en série de l'ECG et du dosage des troponines I

	M +	M –	Total
ECG + et troponine I +	62	1	63
ECG – ou troponine I –	13	24	37
Total	75	25	100

On obtiendrait une spécificité de 0,96 (24/25) et une VPP de 0,98 (62/63). En revanche, la sensibilité, maintenant de 0,83 (62/75), aurait diminué.

En somme, afin de s'assurer de la fiabilité d'un diagnostic en situation d'urgence, il est avantageux de procéder à plusieurs examens simultanément. Si un seul d'entre eux se révèle positif, on considérera le patient comme malade ; il s'agit de la stratégie diagnostique des examens en parallèle. Si, par contre, on dispose de plus de temps, on pourra procéder aux examens de consécutive ; le résultat du dernier examen orientera les décisions au fur et à mesure ; il s'agit de la stratégie diagnostique des examens en série.

À retenir

Les caractéristiques des examens en parallèle et en série

Examens en parallèle	Examens en série
Augmentent la sensibilité	Diminuent la sensibilité
Diminuent la spécificité	Augmentent la spécificité
Augmentent la VPN	Augmentent la VPP
Augmentent les coûts	Limitent les coûts

Lisa discute avec une collègue des mesures de validité intrinsèque de l'étude suisse sur le dosage de d-dimères pour mener à un diagnostic d'embolie pulmonaire. Elles constatent à quel point les résultats de telles études sont importants; en outre, ce type de résultat ne peut provenir de la seule pratique clinique. En effet, elles sont d'avis que dans la pratique clinique courante, on évitera de diriger un patient ayant obtenu un test diagnostique normal vers une procédure de référence efficace et coûteuse. L'information issue de la seule pratique clinique ne permettra donc pas d'établir la proportion de vrais et de faux négatifs. Dans d'autres situations, il est possible qu'il n'existe simplement pas de test de référence pour permettre de calculer ces mesures de validité. Le diagnostic sera alors posé lorsqu'un certain nombre de symptômes ou de signes seront présents. Ainsi, Lisa et sa collègue constatent que lorsque les mesures de validité ne sont pas connues, les décisions relatives aux tests diagnostiques en clinique seront souvent prises sur la foi de données incomplètes ou en comparant deux tests imparfaits. Elles concluent enfin que, dans la pratique clinique, on devra connaître ce qui est susceptible de faire varier la validité d'un examen diagnostique afin de créer les conditions favorables à une performance optimale de cet examen, que les mesures de validité soient connues ou non.

7.5 Les sources de variation dans les examens diagnostiques

Nous avons discuté dans les sections précédentes des différentes manières d'évaluer la validité d'un test diagnostique. Un autre des aspects importants de la performance de ces tests est la reproductibilité. Après avoir traité de la manière de mesurer cette variabilité, nous aborderons les trois grandes catégories de variations susceptibles de modifier les résultats d'un examen: les variations attribuables à l'examineur, les variations attribuables à la personne examinée et les variations attribuables à l'examen lui-même. On qualifie les variations entre observateurs de «différences inter-observateurs». Il existe aussi des différences intra-observateur: une même personne, d'un moment à un autre, peut mesurer ou interpréter différemment ce qu'elle observe. Il s'agit de la «variation intra-observateur».

7.5.1 La mesure de la variation: la mesure d'accord Kappa (K)

Un psychométricien du nom de Cohen a inventé, en se basant sur la notion d'accord, un indice qui reflète la fiabilité de nos instruments de mesure. Il a représenté cet indice, qu'il a nommé «Kappa», par la lettre grecque κ . L'indice Kappa permet d'évaluer globalement la reproductibilité d'un test. La **mesure d'accord Kappa** donne une indication du degré d'accord obtenu au-delà de celui que procure uniquement la chance.

Supposons une situation au cours de laquelle l'accord entre les observateurs est faible, soit inférieur ou égal à la proportion d'accords susceptibles d'être obtenus par hasard; κ serait négatif ou nul. Si, par contre, l'accord entre les observateurs est très

fort, disons tout près de 100 %, et que la probabilité d'un accord dû au hasard est faible, disons de près de 0 %, alors κ serait près de 1.

Ainsi, un κ de près de 1 assure de la reproductibilité de la mesure. En d'autres termes, si plusieurs personnes utilisent une procédure identique dans un tel contexte, elles parviendront sensiblement à la même conclusion; il en sera ainsi d'une seule personne qui mesurera un phénomène à plusieurs reprises. Par contre, un κ de près de 0 ou négatif signifie que l'accord entre les observateurs est bien faible ou que la probabilité d'un accord dû au hasard s'avère très élevée.

Pour la plupart des examens couramment utilisés, les Kappa sont peu élevés. À titre d'exemple, en ce qui concerne la mammographie, le Kappa lié à la mesure inter-observateurs dans le cadre d'une étude menée auprès de 1 214 femmes était de 0,67. Pour l'ECG à l'effort, après avoir procédé à l'examen de 38 tracés (les mêmes), deux cardiologues sont arrivés à un κ de 0,3. Enfin, après avoir examiné les 29 frottis sanguins de femmes souffrant d'anémie par manque de fer, deux pathologistes sont arrivés à un κ de 0,39. Si l'accord intra-observateur est souvent plus élevé, pour les mêmes examens, il se situe autour de 0,6¹¹. On considère généralement que la reproductibilité est bonne lorsque le Kappa excède 75 %.

7.5.2 Les variations attribuables à l'examineur

Les professionnels de la santé doivent prendre des décisions sur la base de données recueillies auprès des malades. Plusieurs facteurs peuvent influencer sur les mesures prises par un examineur; voici les quatre principaux.

D'abord, la perception d'anomalies peut varier d'un examineur à l'autre, simplement en raison de perceptions sensorielles différentes. Par exemple, comme la sensibilité de l'ouïe diffère d'une personne à l'autre, deux personnes n'estimeront pas de la même façon le battement artériel au stéthoscope. De même, selon la sensibilité du toucher, la chaleur d'une articulation ne fera pas nécessairement soupçonner un phénomène inflammatoire de la même manière pour tous.

Par ailleurs, deux observateurs peuvent très bien voir un même phénomène et l'interpréter différemment. Considérons deux personnes qui observent un exsudat dans un œil. Alors que l'une d'elles juge qu'il s'agit d'un exsudat mou, l'autre l'estime dur. Dans ce contexte, puisque ces observateurs voient bien la même image, ce ne sont pas leurs sens qui sont en cause, mais plutôt leur interprétation de la réalité.

La manière de procéder à la classification des observations est un troisième facteur de **variations attribuables à l'examineur**. Par exemple, il arrive souvent que l'examineur ne consigne pas ses observations dans le dossier du patient, mais bien les conclusions qu'il en a tirées. Ainsi, on remarquera souvent au dossier la mention « obésité » plutôt que l'indice de masse corporelle ou le tour de taille. Or, l'obésité peut être définie de plusieurs manières. Dans ce cas, des catégories diagnostiques trop larges peuvent nuire à la précision de la mesure. Il en va de même pour sa validité lorsqu'un diagnostic est défini différemment selon les personnes qui le posent.

Enfin, une autre source de variation tient aux attentes des examinateurs. Il est bien connu que les êtres humains ont tendance à trouver ce qu'ils cherchent. Ainsi, un examinateur qui ignore ce qu'il doit chercher a de fortes chances de ne rien trouver. Par exemple, un étudiant en médecine qui n'a pas encore été sensibilisé à la possibilité de mauvais traitement chez les enfants peut bien mal interpréter l'origine d'une fracture ou d'ecchymoses et passer systématiquement à côté des cas de maltraitance. Réciproquement, un médecin chercheur qui s'intéresse à une maladie rare en trouvera probablement bien plus que ses collègues parce qu'il recherchera systématiquement le moindre indice chez chacun de ses patients. La plupart du temps, l'erreur commise l'est systématiquement et entache par conséquent la validité de la mesure.

7.5.3 Les variations attribuables à la personne examinée

Selon le moment où elles sont recueillies, les valeurs mesurées peuvent différer. L'organisme vivant est en perpétuel mouvement; ainsi, la production hormonale, à elle seule, varie au cours d'une même journée, d'un mois et au fil des années. Considérée isolément, une unique mesure anormale (telle qu'une pression artérielle un peu élevée) ne veut rien dire; il importe de prendre une mesure plusieurs fois avant de conclure à une anomalie.

Aussi, tant la maladie elle-même que les médicaments administrés pour la traiter feront varier les valeurs mesurées. L'anamnèse sera très différente selon qu'elle est recueillie auprès d'un patient dans un moment de délire ou lorsqu'il a tous ses esprits. Le même phénomène se remarquera lorsqu'on interrogera un patient qui a abusé d'une drogue ou d'un médicament. À la limite, un patient interrogé alors qu'il est très souffrant ne fournira pas nécessairement les mêmes indices lorsqu'on l'aura soulagé de sa douleur. Enfin, sachons que la mémoire est intimement liée à des états émotifs, à des sensations et à des contextes particuliers. Ainsi, une mère qui a accouché d'un bébé aux prises avec un problème de santé aura tendance à revivre en pensée toutes les étapes de sa grossesse pour tenter d'expliquer le phénomène, plus qu'une mère ayant donné naissance à un bébé en santé. Les patients atteints d'une maladie grave et évolutive ont la même propension. Ce type de **variation attribuable à la personne examinée** peut donner lieu à un biais de rappel, qui est une erreur systématique d'information.

7.5.4 Les variations attribuables à l'examen

L'environnement influera sur les résultats de plusieurs examens; cela donne lieu à des **variations attribuables à l'examen**. Ainsi, il est vraisemblablement impossible d'obtenir d'une adolescente les mêmes renseignements concernant sa vie sexuelle ou sa consommation de drogues selon que l'examineur se trouve seul avec elle ou que son père assiste à la consultation. La relation entre l'examineur et la personne examinée jouera aussi un rôle non négligeable. Une infirmière dont la charge de travail est lourde n'entendra pas la détresse timidement exprimée par une jeune mère désemparée. De même, la personne âgée bousculée par le médecin s'abstiendra de passer la radiographie pulmonaire qu'il lui a prescrite. La capacité des examinateurs

à utiliser les appareils de mesure est également en cause. La compétence des techniciens de laboratoire, de radiologie, d'échographie, etc. peut grandement influencer sur le résultat des examens et constitue une source de variations dont on doit absolument tenir compte. Qui plus est, un appareil mal calibré ou défectueux transmettra évidemment des mesures invalides.

Le tableau 7.15 donne un aperçu des types de variations — et de leurs sources — susceptibles d'influer sur les résultats d'un examen diagnostique.

TABLEAU 7.15 ■ Les sources des types de variations

Types de variations	Sources de variations
Variations attribuables à l'examineur	<ul style="list-style-type: none"> • Interprétation de la réalité • Classification des diagnostics • Attentes de l'examineur • Lacunes ou insuffisances dans les connaissances
Variations attribuables à la personne examinée	<ul style="list-style-type: none"> • Variations biologiques • Effets de la maladie ou des médicaments • Souvenirs
Variations attribuables à l'examen	<ul style="list-style-type: none"> • Environnement • Relation entre l'examineur et la personne examinée • Utilisation inadéquate des appareils de mesure • Utilisation d'appareils mal calibrés ou défectueux

Source : Traduit et adapté de Sackett, D.L., Haynes, R.B. et Tugwell, P. (1991). *Clinical epidemiology. A basic science for clinical medicine* (2^e éd.), Boston, MA : Little, Brown and Company, p. 36.

En résumé

Dans ce chapitre, les notions de validité intrinsèque (sensibilité et spécificité) d'un examen diagnostique et de validité extrinsèque (valeurs prédictives positive et négative) ont été évoquées. Une bonne validité intrinsèque dépend d'une bonne sensibilité et d'une bonne spécificité, bien qu'il soit généralement impossible d'avoir à la fois une très forte sensibilité et une très forte spécificité; en effet, une augmentation de l'une entraîne généralement une baisse de l'autre. Donc, un compromis guidé par le contexte clinique d'utilisation s'impose. On a noté combien la prévalence de la maladie influe sur la

valeur prédictive des examens diagnostiques. On a constaté que certaines stratégies, telles que la pratique d'examens en parallèle ou en série, permettent de modifier la sensibilité et la spécificité du processus diagnostique et influent sur les valeurs prédictives. On a également vu que des variations intra-observateur et inter-observateurs peuvent survenir et que la mesure d'accord Kappa permet d'en évaluer l'ampleur. Finalement, les variations qui influent sur le résultat de ces examens, soit les variations attribuables à l'examineur, à la personne examinée et à l'examen lui-même, ont été décrites.

Exercices**Voir la section complète des exercices sur le Web.**

1. Vrai ou faux. Un examen doté d'une forte sensibilité est presque toujours positif en présence de la maladie recherchée.
2. Comment désigne-t-on la probabilité que la maladie soit présente si le résultat de l'examen est positif?
 - a) Le seuil de positivité.
 - b) La valeur prédictive positive.
 - c) La sensibilité.
 - d) La spécificité.
 - e) La valeur prédictive négative.
3. On utilise un examen doté d'une forte sensibilité dans chacune des situations suivantes, sauf une. De quelle situation s'agit-il?
 - a) Pour exclure la présence d'une maladie au début d'une investigation.
 - b) Pour réduire le nombre de possibilités diagnostiques.
 - c) Pour diagnostiquer une maladie grave pour laquelle il existe un traitement efficace.
 - d) Pour confirmer un diagnostic.
 - e) Pour dépister une maladie fréquemment observée dans la population.
4. Vrai ou faux. La combinaison d'examens en parallèle permet d'augmenter la sensibilité d'un processus diagnostique.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 8

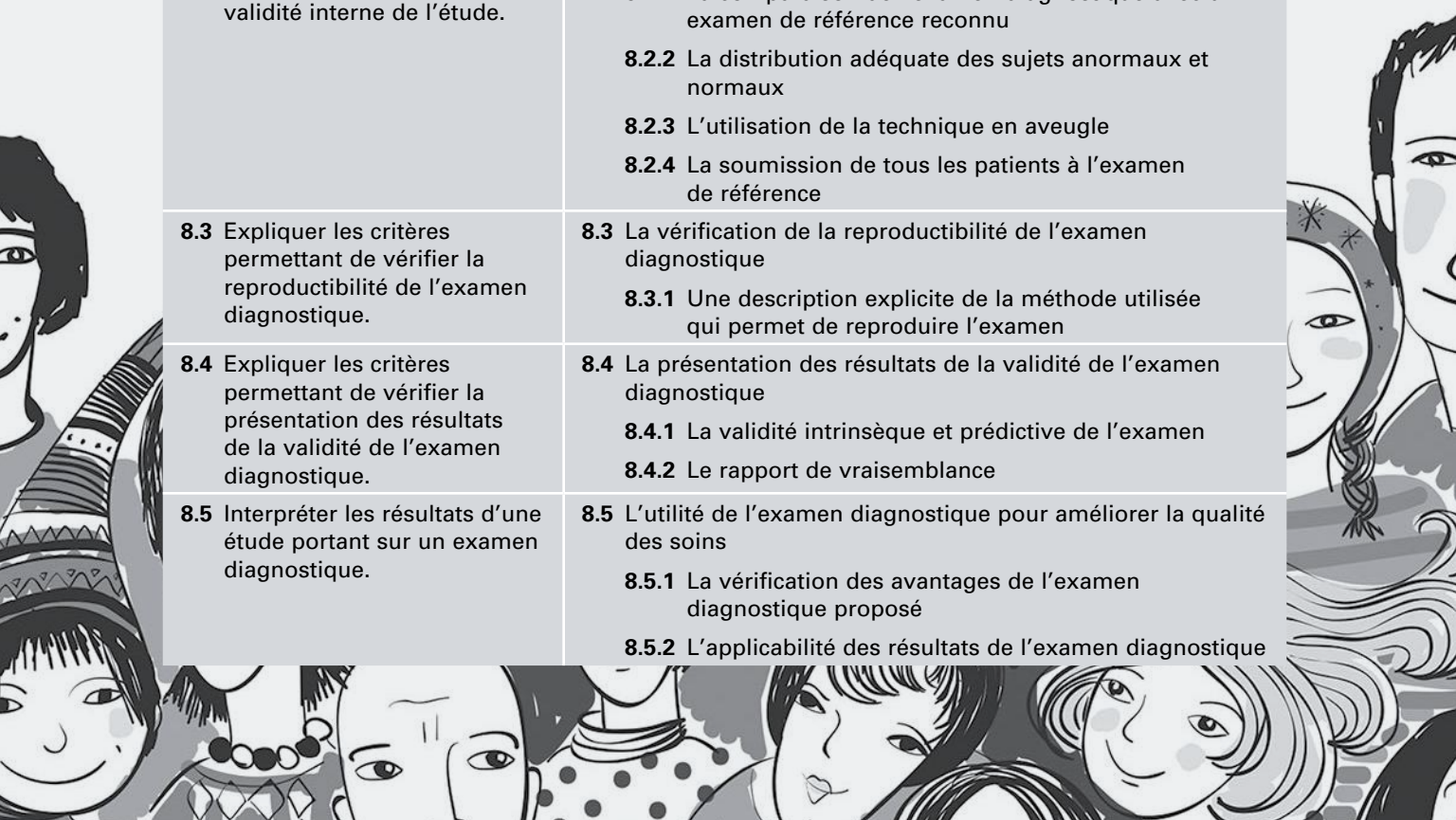
Collaboration à la 3^e édition :
Pamphile Nkogho Mengue
André Simpson

La lecture critique des études portant sur les examens diagnostiques

Au terme de ce chapitre, vous pourrez :

- utiliser un cadre de référence pour la lecture critique des articles scientifiques portant sur des examens diagnostiques ;
- juger de la pertinence de modifier votre pratique à la lumière des résultats d'une publication scientifique portant sur un examen diagnostique.

Objectifs spécifiques	Contenu
8.1 Effectuer systématiquement l'analyse pour juger de la qualité d'une publication scientifique portant sur un examen diagnostique.	8.1 Les critères permettant de juger de la validité d'une étude portant sur un examen diagnostique
8.2 Expliquer les critères permettant de vérifier la validité interne de l'étude.	8.2 La vérification de la validité interne de l'étude 8.2.1 La comparaison de l'examen diagnostique avec un examen de référence reconnu 8.2.2 La distribution adéquate des sujets anormaux et normaux 8.2.3 L'utilisation de la technique en aveugle 8.2.4 La soumission de tous les patients à l'examen de référence
8.3 Expliquer les critères permettant de vérifier la reproductibilité de l'examen diagnostique.	8.3 La vérification de la reproductibilité de l'examen diagnostique 8.3.1 Une description explicite de la méthode utilisée qui permet de reproduire l'examen
8.4 Expliquer les critères permettant de vérifier la présentation des résultats de la validité de l'examen diagnostique.	8.4 La présentation des résultats de la validité de l'examen diagnostique 8.4.1 La validité intrinsèque et prédictive de l'examen 8.4.2 Le rapport de vraisemblance
8.5 Interpréter les résultats d'une étude portant sur un examen diagnostique.	8.5 L'utilité de l'examen diagnostique pour améliorer la qualité des soins 8.5.1 La vérification des avantages de l'examen diagnostique proposé 8.5.2 L'applicabilité des résultats de l'examen diagnostique





Mise en situation

Manon, 24 ans, se présente en début de soirée à l'urgence où l'omnipraticienne Lisa est de garde. Elle ressent une douleur constante au mollet gauche depuis le matin. Manon est en bonne santé et ne présente pas d'antécédents médicaux particuliers. Elle fume depuis l'âge de 17 ans et utilise des anovulants comme moyen contraceptif. À l'examen clinique, Lisa constate une légère enflure au mollet gauche, mais sans chaleur ni rougeur. Manon éprouve une douleur à la palpation du mollet, douleur qui devient plus aiguë lorsque Lisa crée une tension sur le mollet en exerçant une flexion du pied. Elle n'a pas de douleur à la cuisse, mais, à la palpation, elle en ressent une à l'arrière du genou. À la lumière de ces signes cliniques, Lisa ne peut exclure la possibilité qu'une obstruction des veines profondes (phlébothrombose) soit en cause.

Les conséquences d'une phlébothrombose non traitée peuvent être graves, voire fatales, car une embolie pulmonaire peut survenir advenant qu'un caillot (thrombus) se détache et monte vers les poumons. Lisa propose à Manon de subir une échographie Doppler. Elle a récemment lu un article concernant le diagnostic de la phlébothrombose profonde du membre inférieur; on y comparait justement la phlébographie à l'échographie Doppler.

Pertinence clinique

Les cliniciens s'interrogent régulièrement sur la pertinence d'utiliser un examen diagnostique dans une situation clinique donnée. Par exemple, les tests de la fonction pulmonaire pratiqués en série sont-ils utiles dans le suivi des patients asthmatiques? La radiographie des sinus constitue-t-elle un bon examen pour diagnostiquer une sinusite? L'électrocardiogramme est-il indiqué chez les patients qui sont sur le point de subir une intervention chirurgicale?

L'utilisation judicieuse de l'information fournie par les examens diagnostiques exige que le clinicien soit capable d'en apprécier la validité intrinsèque et prédictive, les rapports de vraisemblance, la reproductibilité, ainsi que l'applicabilité clinique. Pour cela, le clinicien doit maîtriser une démarche analytique et systématique qui lui permettra de juger de la qualité et de la validité d'une étude portant sur de tels examens. Cette habileté est d'autant plus importante que de nouvelles technologies font régulièrement leur apparition dans le domaine des soins de santé, ce qui oblige le clinicien à faire preuve d'un solide esprit critique.

8.1

Les critères permettant de juger de la validité d'une étude portant sur un examen diagnostique

Après une recherche documentaire dans les bases de données informatisées, Lisa a retrouvé l'article intitulé *Diagnostic de la phlébothrombose profonde du membre inférieur par échographie Doppler* (voir l'annexe 8.1). S'appuyant sur le cadre de référence pour la

lecture critique d'une publication scientifique (*voir le chapitre 1*), elle juge, après une lecture rapide, que le contenu de l'article est pertinent et que les concepts présentés lui sont accessibles. L'article pourrait lui permettre de juger de la valeur de l'examen auquel elle a invité Manon à se soumettre, ainsi que d'utiliser le résultat de l'examen en s'appuyant sur une connaissance plus complète. Elle décide donc d'en faire une lecture approfondie. Cette lecture lui permettra de vérifier si l'étude est valide sur le plan scientifique. Elle évaluera ainsi la justesse des données et de leur interprétation.

Pour effectuer cette lecture approfondie de façon critique, Lisa décide de s'appuyer sur des grilles d'analyse critique présentes dans la littérature scientifique. Elle retrouve ainsi une grille proposée par Jaeschke¹. Lisa s'appuie également sur une grille de lecture critique conçue en 2007 par Schlienger, Goichot et Meyer², sur une publication de Nendaz³ et, enfin, sur la méthode et la grille STARD 2015^{4 5}. Le groupe STARD est une initiative visant à améliorer la qualité des publications scientifiques portant sur la validité des examens diagnostiques afin de favoriser l'évaluation de la validité interne et externe de ces études. Lisa crée donc la grille de lecture suivante (*voir l'encadré 8.1*) en intégrant, selon quatre grandes questions, les éléments essentiels de ces références.

ENCADRÉ 8.1 ■ Les critères d'analyse d'un article portant sur un examen diagnostique

L'étude présente-t-elle une bonne validité interne ?

- L'examen diagnostique est-il comparé à un examen de référence reconnu ?
- La comparaison des résultats de l'examen diagnostique et de l'examen de référence est-elle effectuée en aveugle ?
- La distribution des anomalies est-elle adéquate ?
- Tous les patients ont-ils été soumis à l'examen de référence ?

L'examen diagnostique possède-t-il une bonne reproductibilité ?

- La méthode utilisée pour effectuer l'examen diagnostique est-elle décrite d'une façon suffisamment explicite pour qu'on puisse la reproduire ?
- La variabilité de l'examen diagnostique est-elle analysée ?
- Les résultats sur la variabilité intra-observateur ou inter-observateurs sont-ils présentés ?

Les résultats de la validité de l'examen diagnostique sont-ils présentés ?

- Quels étaient la sensibilité, la spécificité, les valeurs prédictives et les rapports de vraisemblance de l'examen pour chaque possibilité de résultat ?
- Les résultats de la validité de l'examen sont-ils acceptables dans le contexte clinique précis ?

Cet examen permettra-t-il d'améliorer la qualité des soins offerts aux patients ?

- L'« applicabilité » de l'examen diagnostique est-elle abordée en tenant compte : de la fréquence de la maladie dans d'autres populations, de la faisabilité technique, d'une évaluation économique ?
- Cet examen a-t-il été placé dans le contexte des autres examens utilisés et de leur séquence ?
- Les résultats de l'examen sont-ils applicables à ma patiente Manon ? Changent-ils ma prise en charge ? Profitent-ils à ma patiente ?

Dans les sections qui suivent, nous passons en revue ces critères, en examinant des exemples pour chacun d'eux.

8.2 La vérification de la validité interne de l'étude

8.2.1 La comparaison de l'examen diagnostique avec un examen de référence reconnu

Un examen diagnostique doit être comparé avec l'examen de référence le plus valide possible, c'est-à-dire celui qui permet le plus adéquatement de classer un individu comme étant malade ou non malade. L'examen de référence doit mener à un diagnostic exact et définitif. Il peut s'agir d'une biopsie, d'une chirurgie, d'une autopsie, d'un suivi à long terme ou de toute autre intervention permettant de rendre compte de l'état de santé de la personne examinée avec la plus grande certitude possible. L'examen de référence doit permettre, en dernière analyse, de départager la normalité de l'anormalité.

EXEMPLE 3.1

Dans la section «Méthode» de l'article portant sur l'échographie Doppler, Lisa lit ceci :

L'examen de référence utilisé pour le diagnostic de la phlébothrombose est le défaut de remplissage veineux intraluminal par phlébographie.

Elle se rappelle alors que cet examen consiste à injecter un liquide radio-opaque dans le réseau veineux. L'obstruction d'une ou de plusieurs veines (phlébothrombose) entraînera un défaut de remplissage.

Le critère de comparaison de l'examen diagnostique avec un examen de référence reconnu n'est que partiellement respecté, puisque le seuil à partir duquel le résultat (la phlébographie) est considéré comme anormal n'est pas précisé. Le défaut de remplissage doit-il être complet? Sinon, quel est le degré d'obstruction qui conduit au diagnostic de phlébothrombose? Le fait que le seuil de normalité ne soit pas précisé expose l'étude à un biais d'information. Cela ne permet pas de comparer les résultats de cette étude avec ceux d'autres études similaires. Le manque de détails sur la méthode ne permet pas de l'appliquer de la même manière à d'autres groupes afin d'en vérifier la reproductibilité.

EXEMPLE 3.2

Dans un autre article où le seuil de normalité est précisé, on pourrait lire :

La coronarographie est l'examen de référence pour le diagnostic de la maladie coronarienne. On considère qu'un patient est atteint de la maladie coronarienne lorsqu'on observe un rétrécissement de plus de 90 % de la lumière d'un vaisseau coronarien majeur.

Dans cet article, le seuil de normalité est bien précisé; il devient donc possible de juger si l'examen diagnostique est comparé à un examen de référence reconnu.

8.2.2 La distribution adéquate des sujets anormaux et normaux

Pour qu'il ait une utilité clinique, un examen diagnostique doit permettre de déceler tant les anomalies légères que les anomalies plus graves.

EXEMPLE 8.3

L'article portant sur l'échographie Doppler n'aborde pas directement la question de la distribution des anomalies. Cependant, Lisa lit ce qui suit dans la section « Résultats » :

Par ailleurs, 102 des 103 phlébothromboses situées au-dessus du genou ont été diagnostiquées par échographie comparativement à 12 des 15 phlébothromboses situées au-dessous du genou.

Lisa sait qu'une obstruction veineuse située au-dessus du genou est associée à un plus grand risque d'embolie pulmonaire. Elle peut donc conclure que la validité de l'échographie Doppler a été évaluée chez des patients qui en présentaient des risques plus ou moins élevés.

Dans un même ordre d'idée, il pourra être utile sur le plan clinique de distinguer les résultats normaux et les résultats se situant « dans les limites de la normale ». Les sujets présentant ce dernier résultat pourraient par exemple être revus selon une certaine périodicité.

8.2.3 L'utilisation de la technique en aveugle

Par souci d'objectivité et pour éviter d'introduire un biais d'information, la comparaison des résultats de l'examen diagnostique évalué et de ceux de l'examen de référence doit être effectuée avec la **technique en aveugle**. Les examinateurs qui analysent l'un des examens ne doivent pas connaître les résultats de l'autre examen.

EXEMPLE 8.4

Dans la section « Méthode » de l'article qui l'intéresse, Lisa repère le passage suivant :

Deux groupes distincts de médecins ont interprété et analysé les examens; l'un s'est penché sur les phlébographies, et l'autre, sur les échographies. Les examinateurs ne connaissaient ni les signes cliniques que présentaient les patients ni les résultats de l'analyse de l'autre examen.

Le critère d'utilisation de la technique en aveugle est donc respecté dans cette étude.

EXEMPLE 8.5

Lisa évoque le souvenir de M. Tremblay, un patient vu en consultation l'année précédente. Il toussait alors depuis plus de six mois et ressentait au côté droit une douleur thoracique qui augmentait à chaque inspiration. M. Tremblay était un fumeur; son frère de 51 ans souffrait d'un cancer du poumon.

EXEMPLE 8.5 (suite)

Je me souviens, raconte-t-elle, avoir à l'époque demandé une radiographie des poumons. L'examen n'avait alors rien révélé d'anormal. Peu rassurée toutefois par l'aggravation des symptômes présentés par M. Tremblay, j'avais alors demandé une tomodensitométrie. Les résultats obtenus quelques semaines plus tard révélaient un nodule pulmonaire au segment latéral du lobe moyen. Forte de cette information nouvelle, j'ai examiné à nouveau la radiographie du patient et j'ai cette fois-ci détecté la lésion qui m'avait échappé à la première inspection. Ma connaissance des résultats du second examen a influencé rétrospectivement mon jugement porté sur le premier.

À partir de cette expérience personnelle, Lisa en avait conclu que, dans une étude, il peut se produire un biais d'information notable si les examens diagnostiques et les examens de référence ne sont pas comparés en utilisant une technique en aveugle.

8.2.4 La soumission de tous les patients à l'examen de référence

Le résultat de l'examen diagnostique ne doit pas influencer la décision d'effectuer l'examen de référence. Tous les patients doivent être soumis à l'examen de référence pour éviter un biais d'information que l'on appelle **biais de vérification**. L'évaluation de la validité intrinsèque (sensibilité et spécificité) de l'examen diagnostique serait faussée si les résultats obtenus à cet examen influençaient la décision d'effectuer l'examen de référence.

EXEMPLE 8.6

Dans l'article sur le diagnostic de la phlébothrombose profonde, Lisa prend note du passage suivant, figurant dans la section « Méthode » :

Tous les patients devaient subir une phlébographie et une échographie dans un intervalle maximal de 24 heures.

Le critère de soumission de l'ensemble des patients à l'examen de référence est respecté, et Lisa considère qu'il n'y a pas de biais de vérification.

Un exemple de biais de vérification

Prenons l'extrait suivant :

Cette étude menée auprès de 100 personnes vise à évaluer la validité de l'examen de McBurney pour poser un diagnostic d'appendicite aiguë. L'examen de McBurney consiste à produire, avec la main, une dépression de la paroi abdominale au quadrant inférieur droit, ce qui provoque une douleur s'il y a inflammation de l'appendice. L'examen de référence est la laparotomie. Tous les participants présentant une douleur abdominale aiguë et chez qui l'examen de McBurney s'est révélé positif ont subi une laparotomie, et seulement 40 %

des individus présentant une douleur abdominale aiguë, mais chez qui l'examen de McBurney était négatif ont subi une laparotomie.

Cette étude comporte un biais de vérification évident. En effet, il est possible dans ce cas de départager les individus présentant une appendicite aiguë (vrais positifs) des individus chez qui les résultats de l'examen de McBurney sont positifs, mais qui n'ont pas la maladie (faux positifs). Par ailleurs, on ne possède pas d'information sur l'état de santé réel de 60 % des personnes chez qui l'examen de McBurney s'est révélé négatif. C'est donc dire qu'en ce qui concerne ces sujets, on ignore si la maladie est présente ou non. Cette situation survient parfois dans le cadre d'études où le risque associé à l'examen de référence est tel que le médecin se montre moins enclin à le faire subir aux patients chez qui les résultats de l'examen diagnostique sont négatifs. Cette faille de l'étude conduit à une évaluation erronée de la validité intrinsèque de l'examen diagnostique.

En effet, supposons que, chez 50 personnes, l'examen de McBurney se soit révélé négatif. Cela voudrait dire que 30 ($50 \times 0,6$) personnes n'ont pas subi de laparotomie. Supposons également que ces 30 individus ne souffrent pas d'appendicite aiguë (laparotomie négative). Or, parmi les 20 personnes chez qui l'examen de McBurney s'est révélé négatif et qui ont subi une laparotomie, 8 présentent une appendicite aiguë. Enfin, parmi les personnes chez qui les résultats de l'examen de McBurney étaient positifs, 40 souffrent d'appendicite aiguë. Le tableau de contingence présenté ci-après permet de calculer la sensibilité et la spécificité de l'examen diagnostique.

	Laparotomie +	Laparotomie -	Total
Examen de McBurney +	40	10	50
Examen de McBurney -	8	42	50
Total	48	52	100

Dans ce cas, la sensibilité est de 83,3 %, et la spécificité, de 80,8 %.

Supposons encore que toutes les personnes chez qui les résultats de l'examen de McBurney se sont révélés négatifs et qui n'ont pas subi de laparotomie souffrent d'appendicite aiguë. Le tableau de contingence qui suit résume cette nouvelle situation.

	Laparotomie +	Laparotomie -	Total
Examen de McBurney +	40	10	50
Examen de McBurney -	38	12	50
Total	78	22	100

Cette fois, la sensibilité est de 51,3 %, et la spécificité, de 54,5 %. Il apparaît clairement que le manque d'information sur l'état de santé réel d'un fort pourcentage d'individus chez qui l'examen de McBurney a été négatif entraîne une distorsion de la mesure de la validité intrinsèque de l'examen diagnostique évalué.

8.3 La vérification de la reproductibilité de l'examen diagnostique

La reproductibilité renvoie à la capacité de l'examen à garder une validité semblable quand il est appliqué à diverses populations de patients par des personnes différentes. Cela nécessite deux conditions : d'abord, que la méthode soit décrite avec suffisamment de détails pour pouvoir de nouveau être appliquée de la même manière ; ensuite, que l'interprétation des résultats soit reproductible entre des évaluateurs différents. Ce dernier aspect a été abordé dans le chapitre précédent. Ces mesures permettent de départager l'accord réel de celui qui serait simplement dû au hasard.

8.3.1 Une description explicite de la méthode utilisée qui permet de reproduire l'examen

L'auteur d'un article qui conclut que l'examen diagnostique étudié devrait être utilisé doit indiquer comment réaliser cet examen (la méthode). Cette description doit traiter de la préparation du patient (régime alimentaire, médicaments à prendre ou à éviter, précautions de circonstance à la suite de l'examen), de l'examen lui-même (technique utilisée, possibilité pour le patient examiné de ressentir des effets indésirables) et de la méthode d'analyse et d'interprétation des résultats.

Rassurée sur la validité interne de l'étude, Lisa se propose de vérifier la reproductibilité de l'échographie Doppler.

EXEMPLE 8.7

L'article en question ne comporte pas une description détaillée de la méthode adoptée pour effectuer l'examen diagnostique (*voir l'annexe 8.1*).

Cependant, à la section « Méthode », l'auteur indique qu'il est possible d'obtenir cette information en citant les références précises :

Les techniques utilisées pour la phlébographie et l'échographie ont déjà été décrites par d'autres auteurs.

Par conséquent, en s'appuyant sur de telles références, on peut considérer que l'on satisfait au critère de description de la méthode permettant de reproduire l'examen.

EXEMPLE 8.8

Dans un article fictif, Lisa pourrait lire :

La cholestérolémie totale a été mesurée après un jeûne de 14 heures. Les patients à qui l'on avait prescrit un médicament pouvant exercer des effets sur la cholestérolémie totale ont cessé de prendre ce médicament au moins trois jours avant l'examen. De même, la consommation d'alcool et de caféine a été bannie une semaine avant la prise de sang. L'analyse du prélèvement sanguin a été effectuée selon les méthodes recommandées par la Société canadienne de biochimie. Les résultats de l'examen sont présentés en unités internationales.

La reproductibilité de l'examen

Par ailleurs, Lisa constate que l'article sur le diagnostic de la phlébothrombose profonde ne fournit pas d'information précise sur la reproductibilité de l'examen diagnostique. Tout au plus signale-t-on ceci dans la section « Méthode » :

Deux groupes distincts de médecins ont interprété et analysé les examens; l'un s'est penché sur les phlébographies, et l'autre, sur les échographies. Les examinateurs ne connaissaient ni les signes cliniques que présentaient les patients ni les résultats de l'analyse de l'autre examen. Un premier groupe de trois radiologistes analysait indépendamment les radiographies, et un second groupe regardait l'échographie Doppler sur enregistrement vidéo. Dans les cas où il n'y avait pas d'accord spontané, les deux groupes sont cependant facilement arrivés à un consensus.

Bien que le chercheur ait pris les moyens pour que les radiographies et les échographies soient analysées en aveugle, on ne fournit aucune mesure d'accord dans l'article. Lisa considère donc qu'il lui est difficile de se prononcer sur la reproductibilité de l'examen, malgré le consensus finalement obtenu.

EXEMPLE 8.9

Dans un article encore une fois fictif, on pourrait lire :

Deux psychologues ont évalué en aveugle la présence de symptômes dépressifs chez 40 patients. L'accord entre les examinateurs a été signalé à l'aide de la mesure Kappa, celle-ci étant de 0,54.

Les auteurs de cet article donnent les moyens de juger de la reproductibilité des mesures employées.

8.4 La présentation des résultats de la validité de l'examen diagnostique

Déçue de ne pouvoir juger de la reproductibilité de l'échographie Doppler, Lisa veut tout de même s'assurer que les résultats de la validité de l'examen, c'est-à-dire ses validités intrinsèque et prédictive, sont présentés dans l'article, ou qu'il est possible de les calculer à partir des données qui y sont fournies. Elle remarque alors, à la lecture des articles retenus, que certains auteurs parlent de validité de l'examen, d'autres utilisent le terme « performance de l'examen » ou encore « qualités diagnostiques ». Ces termes constituent des synonymes dans ce contexte.

Les auteurs d'une publication portant sur un examen diagnostique doivent rapporter la sensibilité, la spécificité, les valeurs prédictives et, s'il y a lieu, les mesures de vraisemblance de cet examen.

8.4.1 La validité intrinsèque et prédictive de l'examen

À la section « Résultats », Lisa repère facilement l'information sur la sensibilité globale de l'examen étudié : elle est évaluée à 94 %. Cependant, les auteurs font une distinction quant à la sensibilité de l'échographie selon que l'obstruction veineuse est située

au-dessus du genou (99 %) ou au-dessous de celui-ci (80 %). Quant à sa spécificité, elle est évaluée à 99 %. Les valeurs prédictives ne sont pas précisées, mais Lisa effectue elle-même le calcul à partir d'un tableau de contingence qu'elle construit à l'aide des données de l'étude et qui se présente comme suit.

Il est important de garder en mémoire que les valeurs prédictives pour un examen diagnostique donné dépendent de la prévalence de la maladie. La prévalence est ici de 57 % (118/208). Lorsqu'une étude est réalisée dans un milieu spécialisé, comme dans ce cas à l'urgence d'un grand hôpital, la prévalence de la maladie est souvent très élevée, puisque beaucoup de patients y viennent sur la recommandation d'un médecin. Si une personne présentant une douleur au mollet consulte en clinique de première ligne, la probabilité qu'elle ait une phlébothrombose (prévalence) est beaucoup moins élevée.

	Phlébographies +	Phlébographies -	Total
Échographies +	111	1	112
Échographies -	7	89	96
Total	118	90	208

La valeur prédictive positive est donc dans ce cas de : $\frac{111}{111+1}$, soit 99 %

alors que la valeur prédictive négative est de : $\frac{89}{7+89}$, soit 93 %

8.4.2 Le rapport de vraisemblance

Lisa vérifie également la précision des résultats de la validité présentés. Elle constate qu'aux mesures de sensibilité et de spécificité sont associés des intervalles de confiance dont le niveau de confiance est de 95 %. Lisa juge ces intervalles suffisamment étroits, donc assez précis pour être utiles.

De plus, l'étude fournit de l'information sur les rapports de vraisemblance. Le **rapport de vraisemblance** est un indice qui permet le calcul rapide des valeurs prédictives en fonction de la prévalence de la maladie. Le rapport de vraisemblance possède également d'autres propriétés intéressantes. La définition et l'interprétation du rapport de vraisemblance sont présentées dans l'annexe 8.2.

EXEMPLE 8.10

Dans la section « Résultats » de l'article, Lisa trouve les valeurs des rapports de vraisemblance suivantes, selon la localisation de l'obstruction veineuse :

- Au-dessus du genou, le rapport de vraisemblance positif est de 99, et le rapport de vraisemblance négatif, de 0,01.
- Au-dessous du genou, le rapport de vraisemblance positif est de 80, et le rapport de vraisemblance négatif, de 0,20.

8.5 L'utilité de l'examen diagnostique pour améliorer la qualité des soins

En dernier lieu, Lisa juge l'intérêt clinique de l'examen et évalue s'il peut contribuer à améliorer la qualité des soins offerts à ses patients. En lien avec ses préoccupations, elle s'intéresse donc plus particulièrement aux avantages de l'examen diagnostique et à son applicabilité à ses patients.

8.5.1 La vérification des avantages de l'examen diagnostique proposé

L'examen diagnostique proposé doit offrir certains avantages comparativement à l'examen de référence. Par exemple, il peut comporter moins de risques pour la santé, être plus accessible ou moins coûteux, ou permettre une intervention plus rapide de façon à ralentir ou à freiner l'évolution de la maladie.

EXEMPLE 8.11

Dans l'article sur lequel se penche Lisa, à la section « Introduction », l'auteur souligne les avantages de l'échographie Doppler sur la phlébographie. Il signale son caractère sécuritaire et non effractif ainsi que l'absence de réactions allergiques ou indésirables telles que les bouffées de chaleur ou les nausées, souvent associées à la phlébographie. Il ajoute que l'échographie ne peut provoquer de phlébothrombose profonde ou superficielle, que la grossesse n'est pas une contre-indication à cet examen et que, finalement, celui-ci permet parfois de poser un autre diagnostic (kyste de Baker, anévrisme, thrombophlébite superficielle, altération de la fonction valvulaire veineuse).

Par ailleurs, dans la section « Discussion », l'auteur signale les limites de l'échographie Doppler. Il précise que cet examen, comme la phlébographie, est influencé par l'expertise du radiologiste et que le site anatomique est parfois difficile à déterminer (par exemple, derrière le canal adducteur ou la veine iliaque). Il ajoute que :

[...] l'examen est impossible à pratiquer si le membre est plâtré ou sous traction ou s'il présente des plaies importantes; la visualisation peut être entravée si le patient souffre d'obésité marquée, s'il ne coopère pas, si le membre atteint présente un gonflement important ou si le patient ressent une douleur locale importante.

8.5.2 L'applicabilité des résultats de l'examen diagnostique

Le clinicien doit également évaluer s'il peut obtenir auprès de ses patients des résultats comparables à ceux d'autres patients participant à l'étude. Les caractéristiques démographiques, le niveau de soins (primaire, secondaire ou tertiaire), le type de patients et leur sélection sont des critères importants qui permettent au clinicien de déterminer si sa clientèle et les patients à l'étude présentent un nombre suffisant de traits communs.

EXEMPLE 8.12

À l'urgence, avant d'offrir l'échographie Doppler à Manon, Lisa s'était rappelé le résumé de l'article à la lecture duquel elle avait observé que la population à l'étude provenait d'un contexte de pratique semblable au sien. De plus, à la section « Méthode », elle lit maintenant ceci :

Les patients participant à l'étude ont été sélectionnés parmi ceux qui se sont présentés à l'urgence de notre centre hospitalier avec des symptômes aux membres inférieurs (gonflement, douleur ou rougeur) donnant à penser qu'une phlébothrombose pouvait être en cause. Lorsqu'il recevait un patient, le médecin de l'urgence lui proposait de participer à l'étude et, après avoir obtenu son accord et lui avoir fait signer un formulaire de consentement, le mettait en contact avec le groupe de recherche. Les patients présentant de la fièvre ou une altération de l'état général étaient exclus.

Si le contexte de pratique du lecteur s'apparente à celui qui est rapporté dans l'article, et que ses patients et ceux participant à l'étude présentent un nombre suffisant de traits communs, il peut s'attendre à ce que, dans sa pratique, l'examen diagnostique se révèle d'une efficacité semblable à celle observée dans l'étude qui fait l'objet de l'article.

Lisa constate que le contexte dans lequel les examens ont été réalisés et le type de personnes examinées dans l'étude correspondent de façon générale à ses patients. Elle juge que la validité externe de cette étude est satisfaisante.

L'omnipraticienne en arrive à la conclusion qu'il est dans l'intérêt de Manon de subir une échographie du membre inférieur pour détecter une éventuelle phlébothrombose, étant donné la capacité de cet examen de déceler une telle affection. La sensibilité de l'examen est de 94 %, et sa spécificité atteint 99 %. Avant l'examen, Lisa estime à environ 30 % la probabilité que Manon fasse une phlébothrombose. À partir de ces informations, elle calcule que la valeur prédictive positive est d'environ 99 %, et la valeur prédictive négative, d'environ 96 %. Elle juge donc qu'un résultat positif confirmera le diagnostic ; en revanche, un résultat négatif l'exclura.

Lui ayant fait part de son raisonnement, Lisa propose à Manon sa démarche diagnostique. Si l'examen se révèle positif, Manon devra prendre des anticoagulants ; par contre, si le résultat est négatif, Manon devra être réévaluée dans les deux jours qui suivront, ou plus tôt si les symptômes s'aggravent.

En résumé

Le clinicien s'interroge quotidiennement sur l'utilité des examens diagnostiques. Plusieurs études sont menées pour répondre aux questions qu'il se pose. On a proposé dans le présent chapitre une démarche analytique et systématique qui permet d'évaluer ces études. Leur validité interne

peut être appréciée ; une grille est proposée à cet effet. Leur reproductibilité peut être estimée au moyen des mesures d'accord intra-observateur et inter-observateurs. Les résultats de la validité des examens diagnostiques (validités intrinsèque et prédictive) doivent être présentés dans l'étude ;

si ce n'est pas le cas, il doit à tout le moins être possible de les calculer à partir des données qui y sont fournies. Enfin, avant de pratiquer un exa-

men diagnostique, le clinicien doit s'assurer que l'étude a été réalisée auprès d'une population qui s'apparente à celle de sa pratique.

Exercices



Voir la section complète des exercices sur le Web.

1. Complétez la phrase suivante.

Pendant l'analyse d'un article portant sur un examen diagnostique, le clinicien doit se demander si les résultats de l'étude sont _____ et _____, puis s'ils peuvent contribuer à _____ les soins qu'il donne à ses patients.

2. Un des cinq critères permettant de juger de la validité des résultats d'une étude portant sur un examen diagnostique est manquant dans cette liste. Quel est ce critère?

- A. Il faut comparer l'examen diagnostique avec un examen de référence reconnu.
- B. Une évaluation en aveugle des résultats des examens a été effectuée.

C. La distribution des anomalies est représentative de l'ensemble des cas.

D. ?

E. Une description détaillée de la méthode utilisée permet au lecteur de reproduire et d'interpréter correctement l'examen diagnostique.

3. Vrai ou faux. La prévalence influence le résultat du rapport de vraisemblance.

4. Vrai ou faux. La valeur prédictive positive varie en fonction de la prévalence et du rapport de vraisemblance.

Pour les réponses, voir le corrigé à la fin du manuel.

ANNEXE 8.1

Diagnostic de la phlébothrombose profonde du membre inférieur par échographie Doppler*

Résumé

On a comparé la phlébographie à l'échographie Doppler en mode duplex couleur chez des patients consultant à l'urgence d'un centre hospitalier et manifestant des douleurs pouvant être associées à un diagnostic de phlébothrombose profonde. Les deux examens ont été pratiqués sur un total de 208 membres inférieurs. La sensibilité de l'échographie est évaluée à 99 % (intervalle de confiance [IC] à 95 % de 98,6 à 99,4) lorsque la phlébothrombose se situe au-dessus du genou, et à 80 % (IC à 95 % de 76,8 à 83,2) lorsqu'elle est située au-dessous de celui-ci. Sa spécificité atteint 99 % (IC à 95 % de 99,2 à 99,8). L'échographie Doppler en mode duplex couleur permet d'évaluer les phlébothromboses du membre inférieur d'une façon sécuritaire et non effractive.

* Cet article est fictif. Il a été rédigé par Yvan Leduc, à la suite d'une revue de la littérature sur le sujet.

Introduction

La phlébothrombose profonde du membre inférieur est un problème de santé fréquent et qui échappe souvent au diagnostic. La façon dont elle se manifeste déjoue parfois la vigilance du clinicien. Aux États-Unis, l'incidence est évaluée comme étant approximativement de 1 500 000 cas par année, dont le quart seulement est détecté et traité. Une phlébothrombose non traitée entraîne plusieurs conséquences importantes; ainsi, environ 25 % des patients souffrant d'une phlébothrombose profonde du membre inférieur et qui n'ont pas été traités font une embolie pulmonaire.

Dans sa pratique courante, le clinicien doit fréquemment évaluer des patients présentant des signes cliniques de phlébothrombose. Les facteurs de risque qui prédisposent à ce problème de santé s'observent fréquemment. Parmi ceux-ci figurent l'immobilisation prolongée, la chirurgie et le traumatisme aux membres inférieurs. L'évaluation clinique de la phlébothrombose s'avère difficile, car le questionnaire et l'examen physique sont peu sensibles; en outre, plusieurs maladies se manifestant également par un gonflement, une rougeur ou une douleur peuvent s'apparenter à une phlébothrombose. C'est entre autres le cas de la cellulite, du kyste de Baker, du traumatisme musculaire, de l'insuffisance veineuse, de la phlébite superficielle et de l'insuffisance cardiaque droite. Le diagnostic de phlébothrombose ne pouvant être établi ou exclu sur la seule base de l'examen physique, un examen diagnostique sensible et spécifique doit être effectué.

La phlébographie demeure l'examen de référence, mais, depuis quelques années, un grand nombre d'examens diagnostiques moins effractifs s'offrent comme solutions de rechange. Mentionnons notamment la pléthysmographie par impédance et l'échographie Doppler. On a évalué l'échographie en la comparant à la phlébographie pour le diagnostic de la phlébothrombose profonde. L'échographie Doppler en mode duplex couleur permet la visualisation directe de la veine et du thrombus ainsi que l'évaluation de sa compressibilité. Les avantages potentiels de l'échographie Doppler en mode duplex couleur sur la phlébographie sont les suivants: son caractère sécuritaire et non effractif et l'absence de réactions allergiques ou indésirables telles que les bouffées de chaleur ou les nausées, souvent associées à la phlébographie. De plus, l'échographie ne provoquerait pas de phlébothrombose profonde ou superficielle. L'examen n'est pas contre-indiqué pendant la grossesse, et permet parfois d'établir un autre diagnostic (kyste de Baker, anévrisme, thrombophlébite superficielle, altération de la fonction valvulaire veineuse).

Méthode

Les patients participant à l'étude ont été sélectionnés parmi ceux qui se sont présentés à l'urgence de notre centre hospitalier avec des symptômes aux membres inférieurs (gonflement, douleur ou rougeur) laissant présager qu'une phlébothrombose pouvait être en cause. Lorsqu'il recevait un patient, le médecin de l'urgence lui proposait de participer à l'étude et, après avoir obtenu son accord et lui avoir fait signer un formulaire de consentement, le mettait en contact avec le groupe de recherche. Les patients présentant de la fièvre ou une altération de l'état général étaient exclus. Tous les patients devaient subir une phlébographie et une échographie dans un intervalle maximal de 24 heures.

L'examen de référence utilisé pour le diagnostic de phlébothrombose est le défaut de remplissage veineux intraluminal par phlébographie. Dans le cas de l'échographie Doppler, la présence de l'un des critères suivants permet de porter ce diagnostic : la visualisation du thrombus, l'incompressibilité veineuse et l'absence de flot et de variation du flot avec la respiration. Les auteurs signalent que, tout comme la phlébographie, l'échographie doit être interprétée par un radiologiste. Les techniques utilisées pour la phlébographie et pour l'échographie ont déjà été décrites par d'autres auteurs.

Deux groupes distincts de médecins ont interprété et analysé les examens : l'un s'est penché sur les phlébographies, et l'autre, sur les échographies. Les examinateurs ne connaissaient ni les signes cliniques que présentaient les patients ni les résultats de l'analyse de l'autre examen. Un premier groupe de trois radiologistes analysait indépendamment les radiographies, et un second groupe regardait l'échographie Doppler sur enregistrement vidéo. Dans les cas où il n'y avait pas d'accord spontané, les deux groupes sont cependant facilement arrivés à un consensus.

Résultats

L'étude a été réalisée entre le 1^{er} juillet 2015 et le 31 mars 2016. Le nombre de patients était de 222 au départ, mais 26 patients ont été exclus, étant donné que l'intervalle entre l'échographie et la phlébographie avait dépassé 24 heures pour eux. Parmi les patients, 12 se plaignaient de douleurs aux deux jambes, alors que 208 membres inférieurs ont été étudiés au total.

L'échantillon était composé de 125 hommes et de 71 femmes âgés de 20 ans à 83 ans. Les signes cliniques initiaux étaient les suivants : gonflement (191 membres), douleur (143 membres) et érythème (44 membres). La durée moyenne des symptômes était de neuf jours (étendue : de un jour à trois mois).

On a noté 118 phlébographies évoquant une phlébothrombose comparativement à 90 phlébographies négatives. L'échographie Doppler en mode duplex couleur a permis le diagnostic de 111 phlébothromboses. Par ailleurs, 102 des 103 phlébothromboses situées au-dessus du genou ont été diagnostiquées au moyen de l'échographie comparativement à 12 des 15 phlébothromboses situées au-dessous du genou. Un examen par échographie a conduit au diagnostic de phlébothrombose, alors que la phlébographie était négative.

La sensibilité globale de l'échographie est évaluée à 94 % (IC à 95 % de 92,3 à 95,7). Cependant, la sensibilité de l'examen lorsque l'obstruction veineuse est située au-dessous du genou est de 80 % (IC à 95 % de 76,8 à 83,2), alors que la sensibilité lorsque l'obstruction veineuse se situe au-dessus du genou atteint 99 % (IC à 95 % de 98,6 à 99,4). Sa spécificité totale est évaluée à 99 % (IC à 95 % de 99,2 à 99,8). Pour les obstructions veineuses localisées au-dessus du genou, le rapport de vraisemblance positif est de 99, et celui de vraisemblance négatif, de 0,01. Pour celles localisées au-dessous du genou, le rapport de vraisemblance positif est de 80, et celui de vraisemblance négatif, de 0,20.

Discussion

L'échographie Doppler en mode duplex couleur nous apparaît comme un moyen efficace de diagnostiquer la phlébothrombose profonde. C'est un examen rapide, non effractif et habituellement accessible. Sa sensibilité totale est évaluée à 94 %. Cependant, la sensibilité de l'examen est de 80 % lorsque l'obstruction veineuse se situe au-dessous du genou alors qu'elle est de 99 % lorsque l'obstruction veineuse se situe au-dessus du genou. Sa spécificité atteint 99 %. L'échographie comporte certaines limites : cet examen est influencé par l'expertise du radiologiste, tout comme la phlébographie ; le site anatomique est parfois difficile à évaluer (notamment derrière le canal adducteur ou derrière la veine iliaque) et l'examen devient impossible à pratiquer si le membre est plâtré ou sous traction ou s'il présente des plaies importantes ; la visualisation peut être entravée si le patient souffre d'obésité marquée, s'il ne coopère pas, si le membre atteint présente un gonflement important ou si le patient ressent une douleur locale importante. Dans la présente étude, il y a eu huit patients pour qui les résultats de l'échographie ont pu être altérés à cause de ces limites. Celles-ci peuvent empêcher un diagnostic adéquat et, malgré le fait que les résultats de la présente étude semblent indiquer que l'échographie Doppler offre des avantages marqués, une phlébographie ou des échographies pratiquées en série pourraient être envisagées dans l'éventualité d'un résultat de l'échographie équivoque. En revanche, dans le cas où l'échographie se révèle positive, le diagnostic de phlébothrombose peut être posé, et il n'est pas nécessaire d'effectuer un examen de confirmation avec la phlébographie.

À la lumière de la littérature actuelle et des résultats présentés dans cet article, il est possible d'affirmer que l'échographie Doppler en mode duplex couleur devrait être l'examen diagnostique non effractif de première ligne pour l'évaluation d'une phlébothrombose suspectée aux membres inférieurs. Devant des résultats équivoques ou lorsque le doute clinique demeure, une phlébographie ou des échographies pratiquées en série devraient être envisagées.

ANNEXE 8.2 Le rapport de vraisemblance

Le rapport de vraisemblance (RV) est un indice qui correspond, pour un résultat d'examen donné, au rapport de la probabilité d'être malade sur la probabilité de ne pas l'être. Le rapport de vraisemblance est donc différent selon que le résultat de l'examen diagnostique est positif ou négatif. Cet indice s'exprime simplement, pour un résultat donné, par le rapport entre la proportion d'individus malades (p_{M+}) et la proportion d'individus non malades (p_{M-}), soit :
$$RV = \frac{p_{M+}}{p_{M-}}$$

Exemple d'application

Pour bien comprendre son utilité, calculons le rapport de vraisemblance en utilisant les résultats du dépistage du cancer du col présentés dans le chapitre 7 (voir la section 7.2.3). Selon l'étude de Kang et ses collègues, le seuil de positivité optimal du test careHPV est fixé à 1,74 RLU. Cela signifie que l'examen diagnostique est considéré comme positif lorsque son résultat excède 1,74 RLU, et comme négatif lorsqu'il est égal ou inférieur à ce seuil. À partir des résultats de cette étude et du tableau 7.2, on peut constituer le tableau suivant et calculer le rapport de vraisemblance.

	M +	pM +	M -	pM -
Test careHPV +	120	120/127	1086	1086/9658
Test careHPV -	7	7/127	8572	8572/9658
Total	127	-	9658	-

Pour un seuil de positivité à 1,74 RLU, le résultat et la signification du RV sont :

Pour un résultat positif		Pour un résultat négatif	
Calcul	Signification	Calcul	Signification
$\frac{120/127}{1086/9658} = 8,4$	Un résultat positif survient 8,4 fois plus souvent chez les femmes atteintes d'un cancer du col.	$\frac{7/127}{8572/9658} = 0,062$	Un résultat négatif survient 16 fois moins souvent (1/0,062) chez les femmes atteintes d'un cancer du col.

Une propriété intéressante du rapport de vraisemblance tient au fait qu'on peut le calculer pour n'importe quel seuil de positivité fixé par les chercheurs. Ainsi, il est possible de déterminer ce que serait le rapport de vraisemblance si le seuil de positivité est fixé à 10,0 RLU. Toujours à partir des résultats de l'étude de Kang et ses collègues, et du tableau 7.2, on peut construire le tableau suivant.

	M +	pM +	M -	pM -
Test careHPV +	106	106/127	714	714/9658
Test careHPV -	21	21/127	8944	8944/9658
Total	127	-	9658	-

Pour un seuil de positivité à 10 RLU, le résultat et la signification du RV sont :

Pour un résultat positif		Pour un résultat négatif	
Calcul	Signification	Calcul	Signification
$\frac{106/127}{714/9658} = 11,3$	Un résultat positif survient 11,3 fois plus souvent chez les femmes atteintes d'un cancer du col.	$\frac{21/127}{8944/9658} = 0,179$	Un résultat négatif survient 5,6 fois moins souvent (1/0,179) chez les femmes atteintes d'un cancer du col.

Cet exemple permet d'établir deux constats en ce qui a trait au rapport de vraisemblance.

En premier lieu, le fait que le rapport de vraisemblance soit calculé à l'aide du rapport entre la proportion d'individus malades et d'individus non malades implique que, d'une part, lorsque le résultat de l'examen diagnostique est positif, la proportion d'individus malades correspond à la proportion des vrais positifs (sensibilité), et la proportion des individus non malades, à la proportion des faux positifs (1 – spécificité). En utilisant une sensibilité de 0,9449 et une spécificité de 0,8876 observée avec un seuil de positivité fixé à 1,74 RLU (voir le tableau 7.2), le rapport de vraisemblance peut donc être calculé de la façon suivante :

$$\frac{\text{sensibilité}}{1 - \text{spécificité}} = \frac{0,9449}{1 - 0,8876} = 8,4$$

D'autre part, lorsque le résultat de l'examen diagnostique est négatif, la proportion d'individus malades correspond à la proportion des faux négatifs (1 – sensibilité), et la proportion d'individus non malades correspond à la proportion des vrais négatifs (spécificité). En utilisant une sensibilité de 0,9449 et une spécificité de 0,8876 observée avec un seuil de positivité fixé à 1,74 RLU (voir le tableau 7.2), le rapport de vraisemblance peut donc être calculé de la façon suivante :

$$\frac{1 - \text{sensibilité}}{\text{spécificité}} = \frac{1 - 0,9449}{0,8876} = 0,062$$

La sensibilité et la spécificité d'un examen diagnostique sont des valeurs invariables pour un seuil de positivité donné. Cela sous-entend que le rapport de vraisemblance est également invariable, contrairement à la valeur prédictive qui, elle, varie en fonction de la prévalence.

Le nomogramme de Fagan

Le fait que le rapport de vraisemblance puisse être calculé pour n'importe quel seuil de positivité donné constitue une propriété particulièrement intéressante lorsque le résultat de l'examen diagnostique s'exprime sur une échelle par intervalle. Ainsi, il est possible de refaire chaque fois le calcul du rapport de vraisemblance selon le résultat de l'examen d'un patient.

On peut obtenir, à partir du rapport de vraisemblance, la valeur prédictive positive d'un examen diagnostique si l'on dispose des données de prévalence. Cette opération est réalisable en effectuant quelques manipulations algébriques. Cependant, pour plus de commodité et de rapidité, on peut utiliser le **nomogramme adapté de Fagan**, présenté à la figure A8.2.1.

Ainsi, pour obtenir la valeur prédictive d'un examen diagnostique, on trace une droite qui relie la valeur de la prévalence à celle du rapport de vraisemblance. La valeur prédictive correspond au point de rencontre de cette droite et de l'échelle de la valeur prédictive. Évidemment, on utilise le rapport de vraisemblance approprié en fonction du résultat obtenu.

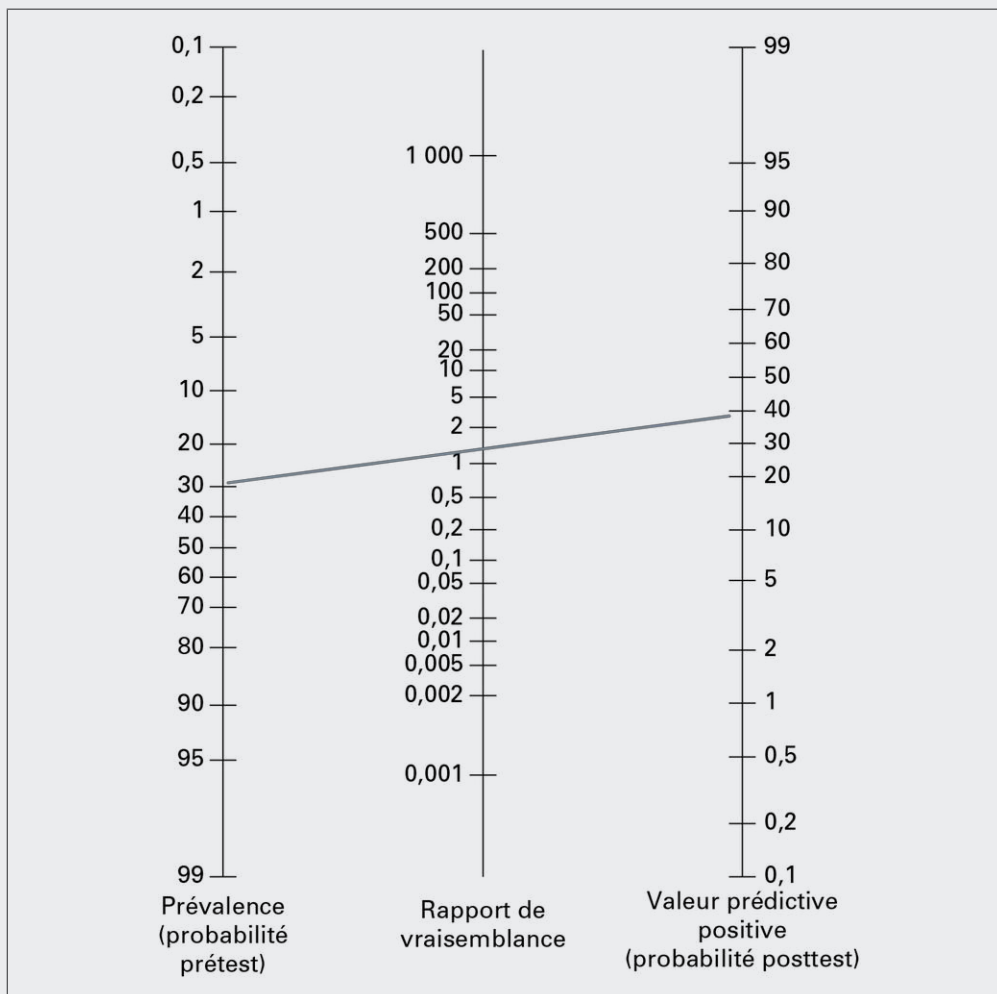
EXEMPLE A8.2.2

Reprenons l'exemple du test du dosage des d-dimères pour le diagnostic de l'emboïlie pulmonaire au chapitre 7, section 7.3. La sensibilité du test est de 97 % et sa spécificité de 40 %, le rapport de vraisemblance ($se / 1 - sp$) est donc de 1,6. En utilisant le nomogramme de Fagan (*voir à la page 171*), à partir de la droite qui relie la valeur de la prévalence (29,0 %) et celle du rapport de vraisemblance (1,6), on obtient rapidement la valeur prédictive positive d'environ 40 %. Le tableau 7.4 (*à la page 739*) permet de vérifier que la valeur prédictive du test est de 39,7 % (281 / 707).

Le rapport de vraisemblance sert également à guider le jugement du clinicien sur l'utilité d'un examen diagnostique. Le tableau suivant présente la signification et l'utilité des différents résultats du rapport de vraisemblance.

Résultat du RV	Signification	Utilité
1	La probabilité d'obtenir un résultat donné (positif ou négatif) est semblable chez les individus malades et chez les individus non malades.	L'examen n'apporte aucun gain d'information et n'est d'aucune utilité.
> 1	La probabilité d'obtenir un résultat positif ou négatif est plus forte chez les individus malades que chez les individus non malades.	L'examen fournit un gain d'information au clinicien et est utile.
< 1	La probabilité d'obtenir un résultat positif ou négatif s'avère moins élevée chez les individus malades que chez les individus non malades.	L'examen fournit un gain d'information au clinicien et est utile.

Lorsque la valeur du rapport de vraisemblance est > 1 ou < 1 , la valeur prédictive positive ou négative diffère de la prévalence, soit de la probabilité de présence de la maladie avant l'examen. Généralement, un examen diagnostique dont le rapport de vraisemblance est plus grand que 10 ou plus petit que 0,1 se révèle utile, étant donné qu'il procure un gain d'information intéressant du point de vue clinique. Par contre, plus le rapport de vraisemblance se rapproche de la valeur 1, moins l'examen est utile, puisque la valeur prédictive tend alors à se rapprocher de la prévalence.

EXEMPLE A8.2.2 (suite)**FIGURE A8.2.1 ■ Un nomogramme permettant d'utiliser le rapport de vraisemblance pour obtenir la valeur prédictive positive d'un examen diagnostique en fonction des données de prévalence**

Source : Adapté de Fagan, T.J. « Letter : Nomogram for Bayes theorem », *New England Journal of Medicine*, 1975, vol. 293, n° 5, p. 257.

CHAPITRE 9

La lecture critique des études descriptives : études de populations

Collaboration à la 3^e édition :

André Simpson

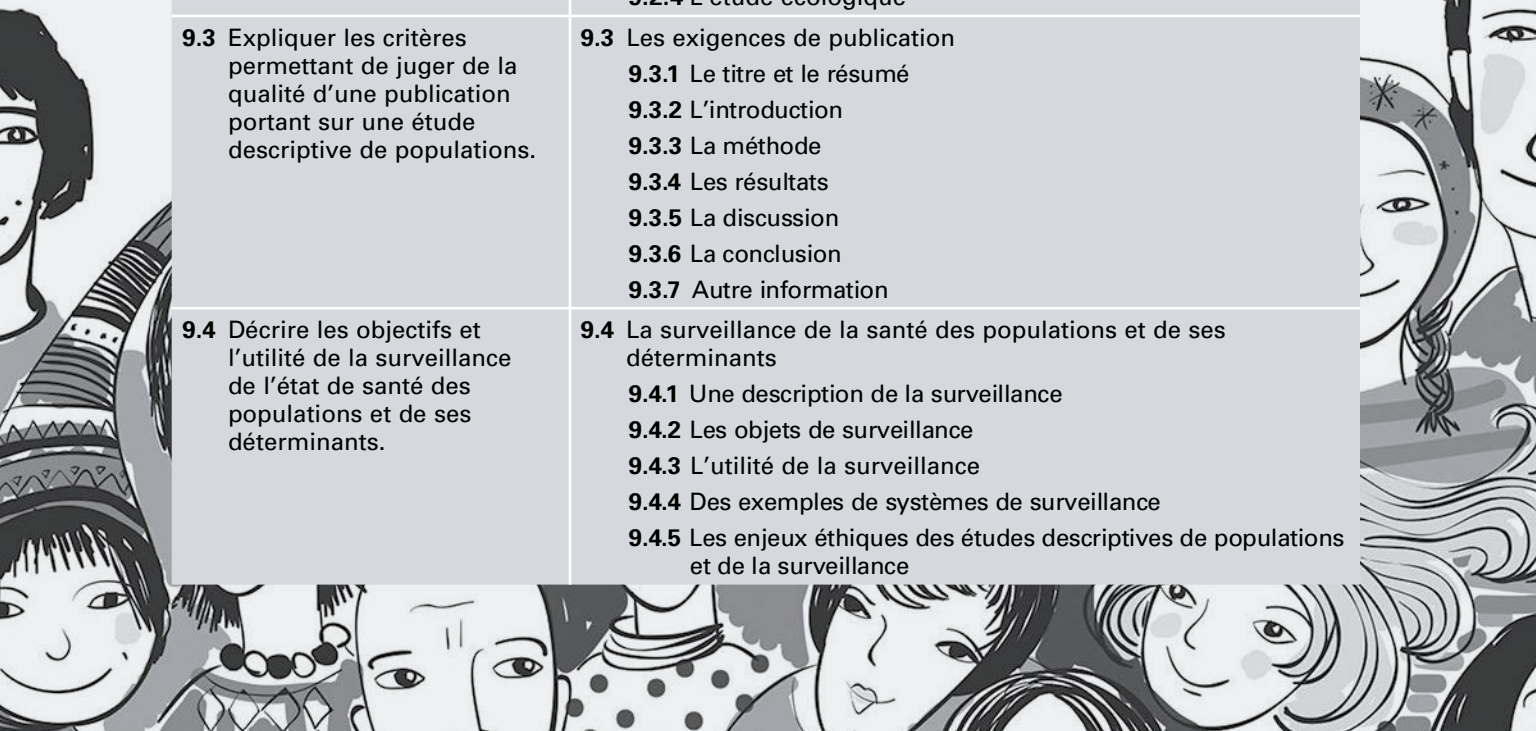
Pierre Deshaies

Thomas Matukala Nkosi

Au terme de ce chapitre, vous pourrez :

- juger de la pertinence d'un article portant sur une étude descriptive de populations ;
- juger de la qualité scientifique d'un article portant sur une étude descriptive de populations ;
- prendre une décision appropriée sur la pertinence de modifier votre pratique clinique à la lumière d'un article portant sur une étude descriptive de populations.

Objectifs spécifiques	Contenu
9.1 Décrire l'intérêt d'une étude descriptive de populations.	9.1 L'utilité des études descriptives de populations 9.1.1 Les avantages d'une étude descriptive de populations 9.1.2 Les limites d'une étude descriptive de populations
9.2 Distinguer les différents types d'études descriptives de populations et établir la pertinence des mesures de fréquence.	9.2 Les types d'études descriptives de populations 9.2.1 L'étude descriptive transversale 9.2.2 L'étude descriptive longitudinale 9.2.3 L'étude de tendance 9.2.4 L'étude écologique
9.3 Expliquer les critères permettant de juger de la qualité d'une publication portant sur une étude descriptive de populations.	9.3 Les exigences de publication 9.3.1 Le titre et le résumé 9.3.2 L'introduction 9.3.3 La méthode 9.3.4 Les résultats 9.3.5 La discussion 9.3.6 La conclusion 9.3.7 Autre information
9.4 Décrire les objectifs et l'utilité de la surveillance de l'état de santé des populations et de ses déterminants.	9.4 La surveillance de la santé des populations et de ses déterminants 9.4.1 Une description de la surveillance 9.4.2 Les objets de surveillance 9.4.3 L'utilité de la surveillance 9.4.4 Des exemples de systèmes de surveillance 9.4.5 Les enjeux éthiques des études descriptives de populations et de la surveillance





Mise en situation

Lisa, l'omnipraticienne, et Jean, le clinicien en santé publique, viennent de prendre connaissance du *Rapport final sur les résultats de la Conférence nationale de concertation sur les maladies évitables par la vaccination au Canada* (mars 2008)¹. Ils se remémorent une période, quelques années auparavant, au cours de laquelle on observait dans la population une recrudescence des cas de méningococcémie, une infection grave disséminée par voie sanguine, causée par une bactérie appelée méningocoque ou *Neisseria meningitidis*. Lisa raconte à Jean le décès par méningococcémie d'un jeune homme de 18 ans qu'elle avait reçu alors qu'elle se trouvait de garde à l'urgence. En parfaite santé à peine 48 heures auparavant, il avait commencé à se sentir fiévreux et « grippé » la veille. Vers midi, il avait senti son état s'aggraver, mais n'avait été transporté à l'urgence qu'à 20 h, en état de choc. Lisa avait alors tout de suite appliqué tous les traitements appropriés, mais sans succès. Elle avait appris par la suite que le jeune homme n'était pas vacciné contre le méningocoque.

Quelques cas de décès d'adolescents et de jeunes adultes causés par cette maladie, qui frappe habituellement davantage les poupons, avaient justement fait les manchettes des journaux les semaines précédentes. Aucun n'était vacciné. Certains médias parlaient d'épidémie et réclamaient une campagne massive de vaccination. Jean relate à son tour que, justement, plusieurs professionnels de la santé publique — dont lui-même — avaient été saisis de ce problème et qu'ils avaient alors évalué la pertinence d'offrir la vaccination à l'ensemble de la population.

Les observations faites à partir d'une série de cas de méningococcémie ont soulevé une

inquiétude du fait que plusieurs personnes décédées présentaient les mêmes caractéristiques : il s'agissait d'adolescents ou de jeunes adultes des quatre coins du Québec et, dans tous les cas, la maladie avait évolué rapidement vers la mort. Ces observations ne permettaient cependant pas de confirmer qu'il y avait une recrudescence des cas de méningococcémie dans la population. Elles ne fournissaient pas non plus d'indications concluantes quant à savoir si ces cas survenaient plus fréquemment parmi les plus jeunes ou les plus âgés, ou en fonction des différentes régions du Québec, des conditions socioéconomiques, etc.

Pertinence clinique

Parce qu'elle tient compte d'un nombre beaucoup plus grand de personnes représentatives de la population, l'étude descriptive de populations permet d'établir la fréquence, la sévérité et la répartition d'une maladie selon des caractéristiques de lieux, de temps et des personnes atteintes dans les populations. Dans le cas présent, une étude descriptive des infections invasives à méningocoques dans la population du Québec s'avérait nécessaire avant de déterminer s'il y avait lieu de vacciner l'ensemble de cette population. Par la suite, pour surveiller l'évolution de cette maladie et l'efficacité des interventions visant à enrayer sa propagation, la mise en place d'un système de surveillance s'est avérée indispensable. Pour ce faire, on a systématiquement recueilli l'information sur la survenue de tous les nouveaux cas afin de calculer le taux d'incidence de la maladie, sa létalité et le taux de mortalité attribuable à cette maladie.

9.1 L'utilité des études descriptives de populations

Dans le chapitre 4, nous avons traité de l'évaluation de la qualité d'une publication portant sur l'étude d'un cas et d'une série de cas. Bien que ces études soient d'un intérêt évident, elles ne permettent pas d'apprécier ou d'estimer l'importance des problèmes de santé étudiés dans la population en général. En revanche, les études descriptives de populations s'attardent à une population composée de sujets atteints et de sujets sains.

L'étude descriptive de populations porte généralement sur des phénomènes spontanés qui surviennent naturellement dans une population donnée. Elle permet de connaître la prévalence ou l'incidence de la maladie étudiée et la mortalité dans une population, selon des variables de personnes, de temps et de lieux. Elle permet aussi de mesurer la fréquence des facteurs de risque et des conséquences d'une maladie selon les mêmes variables.

EXEMPLE 9.1

Un chercheur s'intéresse à la prévalence de l'hypertension artérielle chez les personnes de 18 ans et plus au Québec. Un autre chercheur mesure la mortalité attribuable aux traumatismes routiers depuis 10 ans, dans les différentes régions du Québec et par groupes d'âge. Dans les deux situations, la population du Québec sert de dénominateur pour le calcul des mesures de fréquence de ces problèmes de santé.

9.1.1 Les avantages d'une étude descriptive de populations

Les connaissances acquises par la lecture des études descriptives favorisent une meilleure compréhension de l'**histoire naturelle de la maladie** ou du problème de santé d'intérêt. De telles études permettent de reconnaître certains sous-groupes de la population chez qui la maladie est plus fréquente. De ce fait, elles contribuent à juger de la pertinence ou de l'utilité d'intervenir, au moyen de campagnes de sensibilisation, d'activités de promotion de la santé, de dépistage, de vaccination, etc. Les résultats de ce type d'étude permettent également de formuler des hypothèses étiologiques, c'est-à-dire d'explorer la possibilité qu'il existe des liens de causalité entre des facteurs d'exposition et la maladie ou les décès. Ces hypothèses pourront par la suite être vérifiées notamment par des études à visées étiologiques.

Puisque le vaccin contre le méningocoque disponible au début des années 2000 pouvait comporter certains inconvénients, Jean et Lisa se rappellent qu'avant de l'offrir à tous, ils avaient dû s'assurer que le risque de méningococcémie était suffisamment important chez certains groupes de la population.

Ainsi, après avoir observé et décrit les quelques cas de décès par maladie invasive à méningocoques, Jean avait étudié la fréquence et la répartition de cette maladie dans la population. De plus, il s'était efforcé de comprendre l'évolution de ce phénomène dans le temps de manière à déterminer si l'augmentation soupçonnée du nombre de cas était bien réelle. Une étude descriptive des infections à méningocoques pour les 10 dernières années avait alors été réalisée.

Les résultats de l'étude descriptive de populations permettent également de vérifier la pertinence de certains programmes ou services de santé visant à prévenir et à contrôler la propagation de la maladie dans la population. Ce type d'étude peut aussi servir à décrire les effets secondaires des médicaments et les répercussions de certaines interventions, qu'il s'agisse de traitements, de campagnes de vaccination, de programmes de promotion de la santé ou de dépistage de problèmes de santé, etc. Enfin, en santé publique, la connaissance des groupes à risque permet de mieux orienter les interventions et les moyens de prévention.

EXEMPLE 9.2

Au début des années 2000, les études descriptives réalisées dans les pays industrialisés ont permis d'établir que l'épidémicité des infections invasives à méningocoques se traduit par trois changements principaux, soit une létalité à la hausse, un déplacement de la répartition des cas vers les groupes plus âgés et l'apparition d'une souche bactérienne prédominante plus virulente. L'étude descriptive a également permis de cerner l'évolution de la létalité et de la répartition des cas selon l'âge. Ces données, combinées à la connaissance des autres caractéristiques des personnes atteintes, ont conduit à une décision éclairée quant à la pertinence de réaliser une campagne massive de vaccination contre le méningocoque.

9.1.2 Les limites d'une étude descriptive de populations

La principale limite de ce type d'étude tient à l'absence de groupes de comparaison. En effet, bien qu'elle permette de dégager certaines connaissances utiles à la formulation d'hypothèses étiologiques, cette limite fait en sorte qu'une telle étude ne peut servir à vérifier l'existence de liens de causalité entre un facteur donné et une maladie.

Comme dans la plupart des types d'études, différents biais peuvent entacher la validité. Des biais de sélection, par exemple dans la définition de cas utilisée, peuvent entraîner une sous-estimation ou une surestimation des mesures de fréquence de la maladie et des décès.

Pour diminuer la possibilité de biais de sélection, on devra, dans la mesure du possible, recourir à des stratégies d'échantillonnage aléatoire.

Des biais d'information peuvent aussi influencer sur les résultats d'une étude descriptive de populations. Ainsi, un biais de rappel peut notamment survenir dans les enquêtes effectuées au moyen de questionnaires qui font appel à la mémoire. Des biais de confusion peuvent survenir lorsqu'on compare la fréquence d'un problème de santé entre divers sous-groupes de la population ayant des caractéristiques différentes, par exemple d'âge, de sexe ou de lieu. L'**analyse stratifiée** et le recours aux méthodes d'ajustement et de standardisation des mesures de fréquence permettront de contrôler l'effet d'un biais de confusion et de réaliser une comparaison valide de ces résultats.

Les études descriptives de populations peuvent également mener à une estimation imprécise des mesures de fréquence lorsque la maladie ou l'événement étudié est rare ou lorsque le nombre d'individus qui compose l'échantillon étudié est insuffisant. Ainsi, avant de généraliser les résultats de ces études à l'ensemble de la population cible, on devra tenir compte de la précision de l'étude.

EXEMPLE 9.3

Plusieurs sources potentielles de données s'offraient aux épidémiologistes pour décrire les infections à méningocoques. Pour connaître le nombre de cas, ils auraient pu consulter des sources de données existantes, tels le fichier des maladies à déclaration obligatoire et les registres ou les fichiers de données hospitalières. Des données inédites auraient également pu être obtenues en procédant à une enquête auprès des autorités locales de santé publique, des hôpitaux et des médecins. Cette deuxième stratégie était toutefois plus laborieuse et onéreuse. Par ailleurs, pour obtenir des données de population servant de dénominateur, Jean et ses collègues ont pu s'appuyer sur les données officielles de recensement correspondant à la région qui les intéressait.

L'informatisation en constante évolution des dossiers des patients dans les centres hospitaliers et les milieux cliniques constitue une opportunité nouvelle pour des recherches épidémiologiques. On doit alors s'assurer de respecter les exigences légales et éthiques d'accès à de telles informations.

L'exploitation des différentes sources de données, de même que certaines stratégies de collecte d'information nécessaires à la réalisation d'une étude descriptive et à la lecture critique des articles scientifiques relatifs à ce type d'étude, pose le défi de l'adéquation entre les stratégies d'analyse des données et des caractéristiques du problème de santé à l'étude, comme en témoigne l'exemple suivant.

EXEMPLE 9.4

Pour évaluer la recrudescence réelle des cas de maladie invasive à méningocoques durant les 10 années antérieures, Jean aurait pu se baser uniquement sur les données issues du fichier national des maladies à déclaration obligatoire. Or, historiquement, il a été observé que les médecins et les laboratoires ont omis de déclarer des cas d'infection à méningocoques certaines années, entraînant une sous-estimation du nombre réel de cas pour ces années. Par ailleurs, le battage médiatique suscité par la recrudescence de cette maladie aurait très bien pu inciter les médecins à déclarer davantage ces cas. L'augmentation de la mesure d'incidence de la maladie aurait plutôt traduit, dans ce contexte, une hausse de la déclaration. Conclure à un accroissement réel du nombre de cas aurait été une erreur.

Pour éviter cette distorsion, Jean a complété les données tirées du fichier national des maladies à déclaration obligatoire avec celles du fichier national sur les hospitalisations. Il est en effet légitime de croire que la grande majorité des patients souffrant de maladie invasive à méningocoques se sont présentés à un hôpital de soins aigus pour recevoir un traitement. Cette source de données, vraisemblablement plus exhaustive, minimise un biais d'information lié aux pratiques, variables dans le temps, de déclaration des cas par les médecins.

Les variables de personnes que l'on considère le plus souvent dans les études descriptives sont l'âge, le sexe, l'appartenance ethnique, les conditions socioéconomiques (c'est-à-dire le revenu, le niveau de scolarité et le statut d'emploi), l'environnement

familial et social ; sont également considérées les variables pertinentes au sujet d'intérêt telles les habitudes de vie (la sédentarité, la consommation de tabac ou d'alcool, etc.) ou l'exposition à des contaminants en milieu de travail.

Les variables de temps permettent de décrire des variations cycliques (comme les variations saisonnières, annuelles ou décennales) des problèmes de santé et de leurs déterminants.

EXEMPLE 9.5

Le fichier de données sur les hospitalisations contient plusieurs informations retenues pour l'étude, notamment l'âge et le sexe du patient, sa date d'entrée et de sortie de l'hôpital, les diagnostics précis et, s'il y a lieu, le décès en cours d'hospitalisation.

Les variables de lieux peuvent concerner des ensembles très différents allant des grands regroupements géographiques séparés par des barrières naturelles (chaîne de montagnes, cours d'eau, etc.) aux divisions administratives (pays, provinces et régions sociosanitaires), en passant par les zones urbaines et rurales, et jusqu'aux microenvironnements comme les classes d'une école ou les aires de travail dans une entreprise.

EXEMPLE 9.6

Compte tenu de l'état des connaissances sur les infections à méningocoques, Jean avait souhaité obtenir des données sur le lieu exact où chaque individu avait contracté la maladie, ainsi que sur certaines caractéristiques particulières de son milieu de vie (nombre de personnes vivant sous le même toit, tabagisme, etc.). Cependant, pour avoir accès à ces données, il lui aurait fallu joindre chaque médecin ou chaque patient survivant, car le fichier de données sur les hospitalisations ne mentionne ni la municipalité de résidence du patient ni l'hôpital où il s'est présenté dans un premier temps. Une solution possible : consulter chaque dossier hospitalier, ce qui représente un travail énorme. Heureusement, une équipe d'archivistes médicaux a appuyé Jean dans cette tâche.

9.2 Les types d'études descriptives de populations

9.2.1 L'étude descriptive transversale

Lorsqu'on réalise une étude descriptive transversale, on s'intéresse à un événement, à une caractéristique ou à une maladie à un moment précis dans le temps ; par exemple, au nombre ou à la proportion d'individus atteints d'une maladie à un moment donné. Il s'agit d'un portrait instantané de la situation. On y mesure le nombre de cas prévalents, à savoir le nombre existant de malades dans la population, et la prévalence de la maladie, soit la proportion des personnes atteintes de cette maladie (*voir à ce sujet le chapitre 3*). Une telle étude est particulièrement utile pour déterminer la fréquence et la répartition des problèmes de santé touchant la population à un moment précis.

EXEMPLE 9.7

Lors de l'enquête portant sur la santé des collectivités canadiennes en 2003, on a mesuré les proportions de personnes âgées de 12 ans et plus ayant déclaré une maladie chronique diagnostiquée par un professionnel de la santé. Au Québec, 24,9 % des répondants ont déclaré être atteints d'allergies non alimentaires, 14,0 % d'arthrite et 14,5 % d'hypertension comparativement à respectivement 28,1 %, 17,7 % et 14,4 % pour le reste du Canada. Les prévalences observées pour le Québec au cours d'une enquête subséquente en 2005 étaient similaires à celles observées en 2003².

9.2.2 L'étude descriptive longitudinale

Le chercheur qui entreprend une étude longitudinale s'intéresse à l'apparition des nouveaux cas d'une maladie durant une période donnée (appelés aussi **cas incidents**). Pour réaliser ce type d'étude, il peut observer la population totale. Mais généralement, il observera un échantillon, un sous-ensemble de cette population constitué d'une cohorte d'individus représentative de la population totale.

Le plus simple pour un chercheur est de travailler dans le contexte d'une population dynamique fermée. Comme on l'a vu dans le chapitre 3, on connaît bien chaque individu qui constitue la population dans un tel contexte. La cohorte sera formée au temps zéro de personnes qui ne sont pas atteintes de la maladie. Elles seront observées jusqu'à l'apparition de la maladie, jusqu'à leur décès ou jusqu'à la fin de l'étude.

Mais cela n'est pas toujours possible. Le chercheur devra alors travailler dans le contexte d'une population dynamique ouverte. Par exemple, sa cohorte pourrait être constituée des habitants d'une ville. Il devra alors tenir compte qu'au temps zéro, un certain nombre de personnes auront déjà la maladie, que des gens emménageront dans la ville, que d'autres la quitteront, qu'il y aura des naissances et des décès, etc. Il lui faudra alors utiliser une série de techniques pour tenir compte de ces variations.

Ces études permettront d'établir les mesures de fréquence suivantes: l'incidence, l'incidence cumulée, le taux d'incidence, la létalité, la survie et le taux de mortalité pour la durée d'observation.

EXEMPLE 9.8

Les données du fichier des hospitalisations et du fichier des maladies à déclaration obligatoire qui couvrent la population totale du Québec n'avaient pas permis l'identification des personnes en bonne santé au début de la période d'observation. Jean avait donc dû travailler dans le contexte d'une population dynamique ouverte. Il avait ainsi comptabilisé 799 nouveaux cas d'infection invasive à méningocoques entre le 1^{er} avril 1981 et le 31 mars 1991. Pour calculer le taux d'incidence, il avait utilisé au dénominateur la somme de la population québécoise moyenne pour les 10 années à l'étude, soit 67 447 257 personnes-année. Ces données avaient permis d'estimer un taux d'incidence d'infection à méningocoques de 1,2 nouveau cas pour 100 000 personnes-année (soit $799/67\,447\,257 \times 100\,000$).

EXEMPLE 9.8 (suite)

En plaçant au numérateur le nombre de décès attribuables à la maladie survenus en cours d'hospitalisation, soit 43 décès, Jean avait établi le taux de mortalité attribuable à cette maladie à 0,064 décès pour 100 000 personnes-année, ou, exprimé autrement, à 6,4 décès pour 10 000 000 personnes-année.

À partir des 799 nouveaux cas d'infection invasive à méningocoques qu'il a inventoriés, Jean avait également mesuré le temps écoulé entre le début du traitement et le décès pour chacun des cas.

Cette mesure avait permis à Jean de connaître la durée et la probabilité de survie associées à cette maladie parmi les cas traités. Il avait aussi pu calculer la létalité des infections invasives à méningocoques de la façon suivante : 43 décès sur 799 cas entre le 1^{er} avril 1981 et le 31 mars 1991, soit 5,4% pour une durée d'observation de 10 ans.

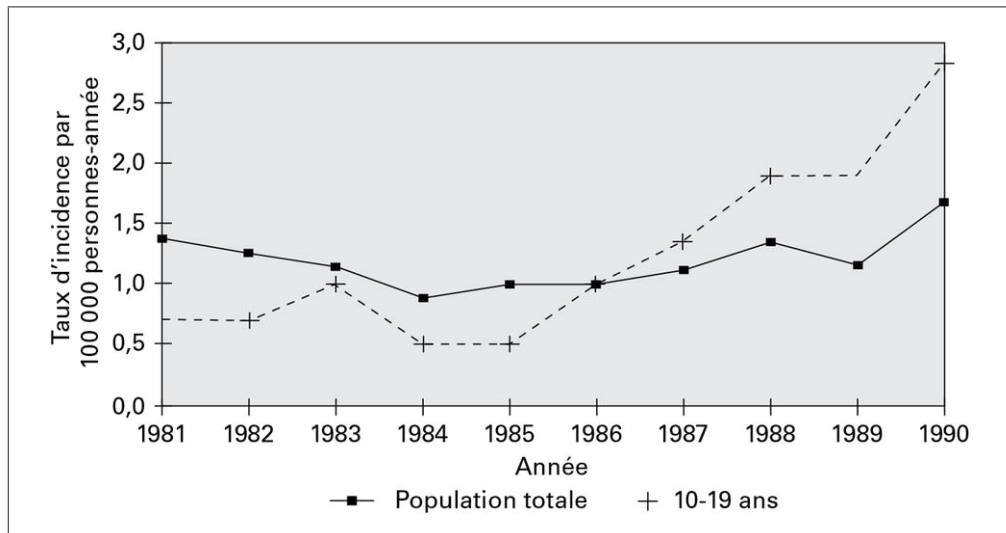
En repensant à tout cela, Jean et Lisa se disent satisfaits d'avoir contribué, par leur action et par la réalisation de cette étude descriptive de la fréquence et de la répartition des infections invasives à méningocoques, à l'établissement des connaissances de base qui ont permis de formuler les recommandations pour la vaccination de populations ciblées.

9.2.3 L'étude de tendance

L'étude de tendance vise à mesurer la fréquence et à évaluer l'évolution d'un déterminant, d'une maladie ou des décès dans le temps. Le chercheur répète alors une même étude à des périodes ou à des moments différents, ou il sépare simplement ces données en périodes d'intérêt. L'étude de tendance est réalisée à partir de données de prévalence, d'incidence ou de mortalité. Elle nécessite que l'on collecte les données à au moins trois moments distincts ou que l'on considère trois épisodes différents. De telles études peuvent couvrir une longue période. Une étude de tendance permettrait de connaître, notamment, l'évolution d'un problème en lien avec des interventions tels des programmes de prévention ou de dépistage dans la population.

EXEMPLE 9.9

Jean avait étudié l'évolution du taux d'incidence des infections invasives à méningocoques dans la population totale et dans chaque groupe d'âge d'année en année pour les 10 ans à l'étude comme le présente la figure 9.1. Les taux d'incidence annuelle et leur évolution avaient montré que, depuis 1986, l'augmentation était plus marquée dans le groupe des 10-19 ans que dans la population générale. Cette information, combinée à une évaluation de la pertinence d'une vaccination systématique, avait permis à Lisa de mieux soupeser le bénéfice réel que ses jeunes patients retireraient si l'ensemble des personnes du groupe d'âge auquel ils appartiennent recevait le vaccin.

EXEMPLE 9.9 (suite)**FIGURE 9.1 ■ Le taux d'incidence annuelle des infections invasives à méningocoques dans la population totale et dans le groupe des 10-19 ans de 1981 à 1990 au Québec**

Source: Adapté de Deshaies, P. «Épidémiologie des infections invasives à *Neisseria meningitidis* au Québec de 1981 à 1990», mémoire de maîtrise en épidémiologie, Québec, Québec, Université Laval, 1993, p. 34-35.

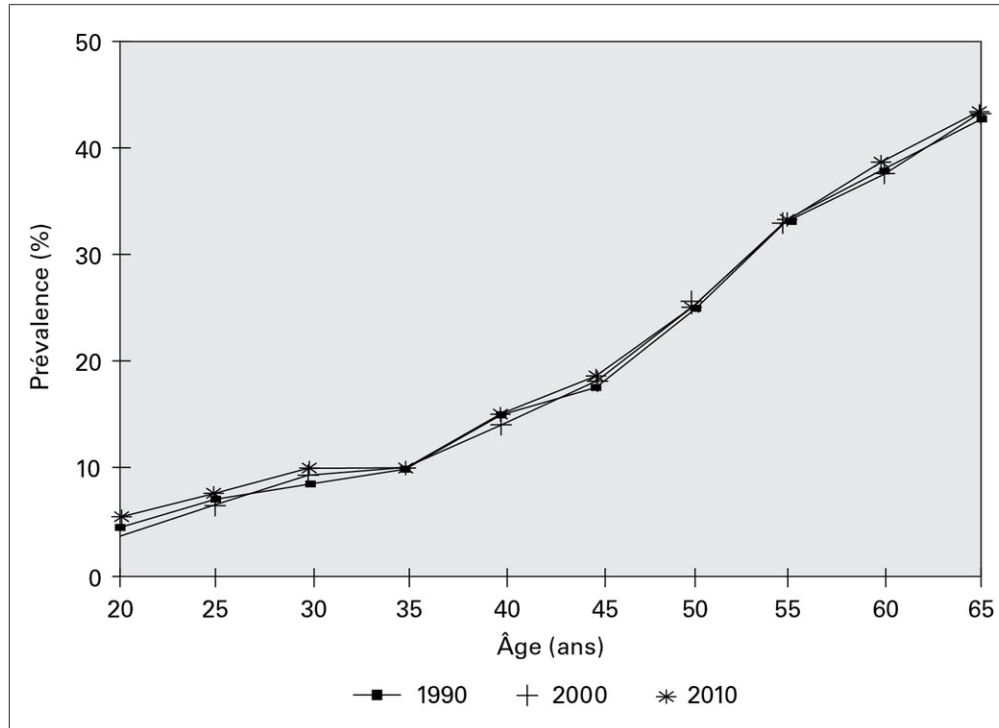
Les effets de l'âge, de période et de cohorte

L'analyse et l'interprétation des résultats d'une étude de tendance visent généralement à mesurer trois effets liés au temps écoulé durant la période totale: l'effet de l'âge, l'effet de période et l'effet de cohorte³.

L'effet de l'âge existe lorsque la fréquence de la maladie varie avec l'âge, indépendamment de la période étudiée et de l'année de naissance des individus.

EXEMPLE 9.10

Une étude de tendance visant à connaître l'évolution de la prévalence de la perte auditive est réalisée. On considère qu'un sujet présente un problème d'audition important lorsque la perte moyenne d'audition mesurée aux fréquences 0,5 kHz, 1 kHz, 2 kHz et 4 kHz à l'audiométrie est de 25 dB ou plus. L'étude de la prévalence de la surdit  est r alis e chez des adultes  g s de 20   65 ans en 1990, 2000 et 2010. On observe que la baisse de la capacit  auditive due   l' ge (presbyacousie) affecte tous les individus de la population et que la proportion de la population pr sentant une telle baisse augmente avec l' ge, par le simple effet du vieillissement normal. Les courbes de pr valence des atteintes auditives pr sent es   la figure 9.2 illustrent ce ph nom ne. Quelle que soit l'ann e de l' tude, et donc l'ann e de naissance des sujets, on trouve la m me pr valence pour un  ge donn .

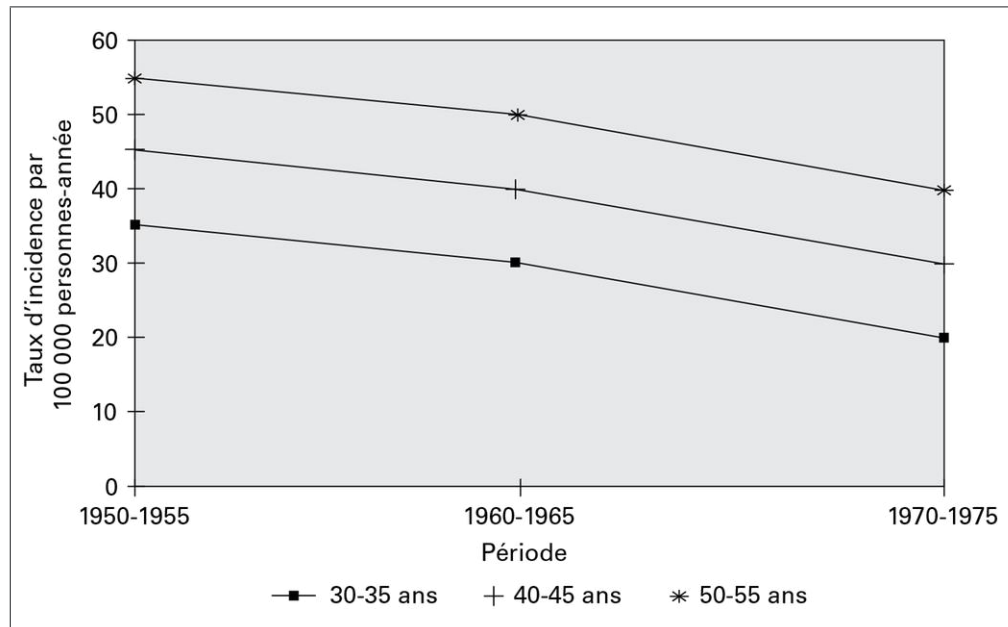
EXEMPLE 9.10 (suite)**FIGURE 9.2** ■ La prévalence des atteintes auditives selon l'âge en 1990, 2000 et 2010 (données fictives)

L'effet de période existe lorsque la fréquence de la maladie varie avec le temps, indépendamment de l'âge des individus et de leur année de naissance.

EXEMPLE 9.11

Une étude de tendance vise à comparer l'incidence de la silicose chez les travailleurs miniers du Québec à diverses époques. On compare les résultats des données collectées au cours des périodes 1950-1955, 1960-1965 et 1970-1975. On observe, comme l'indique la figure 9.3, à la page suivante, une incidence plus grande au cours de la période 1950-1955 pour toutes les catégories d'âge.

Deux facteurs favorisant l'apparition de la silicose durant les années 1950 expliquent cette situation : d'une part, la tuberculose, associée à l'exposition aux poussières minérales, pourrait accélérer le développement de la silicose ; d'autre part, le manque de ventilation dans les mines occasionnerait une plus forte concentration de poussières minérales. Cette dernière cause a été réduite par une meilleure ventilation. Par ailleurs, la fréquence de la tuberculose pulmonaire a beaucoup diminué depuis cette époque.

EXEMPLE 9.11 (suite)**FIGURE 9.3 ■ L'incidence de la silicose chez les travailleurs miniers âgés de 30 à 55 ans au cours des périodes 1950-1955, 1960-1965 et 1970-1975 (données fictives)**

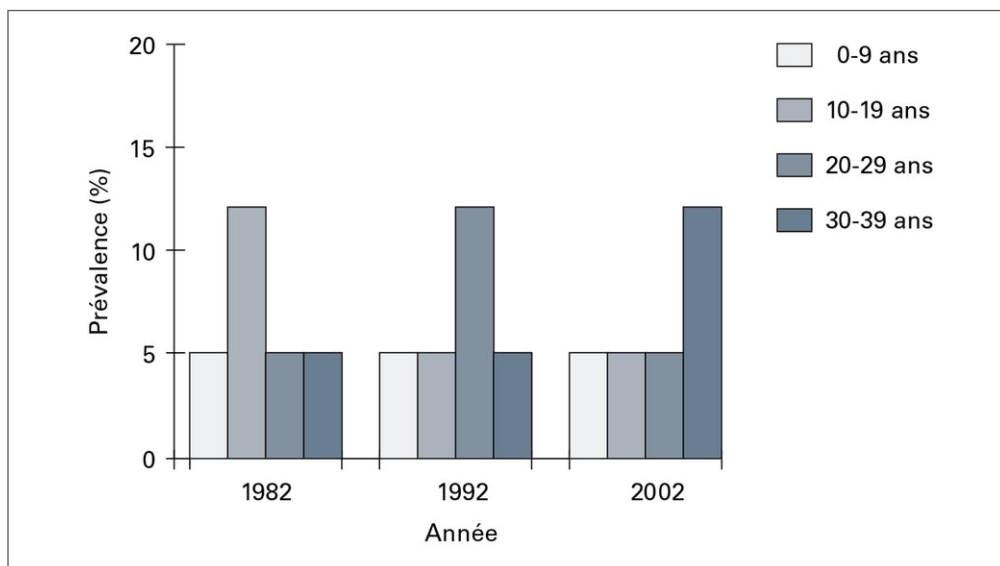
L'effet de cohorte existe lorsque la fréquence de la maladie varie en fonction de l'année de naissance des individus, indépendamment de leur âge.

Une tendance observée, à la hausse ou à la baisse, peut ainsi résulter de l'un de ces trois effets ou d'une combinaison de ceux-ci. La connaissance de facteurs étiologiques particuliers permet parfois de les isoler, comme l'illustrent les exemples précédents. Dans d'autres situations, il peut être difficile de distinguer leur effet respectif.

EXEMPLE 9.12

Les auteurs d'une étude de tendance s'intéressent à la prévalence de la surdité congénitale dans la population d'Israël entre 1982 et 2002. Trois mesures de prévalence ont été réalisées, soit en 1982, en 1992 et en 2002. Les résultats présentés par groupes d'âge à la figure 9.4 laissent voir une prévalence significativement plus élevée chez les 10-19 ans en 1982, chez les 20-29 ans en 1992 et chez les 30-39 ans en 2002.

Ce phénomène s'explique par une **épidémie** importante de rubéole dans la population d'Israël en 1972. Une grande proportion des femmes enceintes a alors été atteinte. Or, on sait que la rubéole chez la femme enceinte constitue une cause importante de surdité congénitale. La prévalence plus élevée est observée dans la cohorte des individus nés durant l'année 1972.

EXEMPLE 9.12 (suite)**FIGURE 9.4 ■ La prévalence de la surdité congénitale dans la population d'Israël en 1982, 1992 et 2002 (données fictives)**

9.2.4 L'étude écologique

L'étude écologique, qualifiée aussi d'étude corrélationnelle, est un type particulier d'étude descriptive de populations. Sa spécificité réside dans le fait que l'unité d'observation pour les variables de personnes, de temps et de lieux n'est plus l'individu lui-même avec ses caractéristiques individuelles (âge, sexe, lieu de résidence), mais un groupe d'individus, tels une classe, les travailleurs d'une usine ou les habitants d'un quartier, d'une ville, d'une région, voire d'un pays. Ainsi, comme pour les autres types d'études, on doit disposer d'information sur le problème de santé à l'étude et sur le ou les facteurs d'intérêt qui y sont potentiellement associés pour chacune des unités d'observation.

L'étude écologique peut donc prendre la forme d'un des trois types d'études décrits précédemment. Les données sur l'exposition prennent alors la forme d'un indicateur de groupe : par exemple, la consommation d'alcool estimée à partir des taxes perçues ou l'exposition à l'arsenic contenu dans l'eau potable à partir des analyses d'eau potable par réseau d'aqueduc. On comprend que l'on n'obtient pas, dans le cadre de ce type d'étude, d'information sur la consommation individuelle d'alcool ou d'arsenic, mais bien sur la consommation collective, puisque l'on ne tient pas compte des caractéristiques de chaque individu, mais plutôt de celles de son groupe d'appartenance.

Ce type d'étude présente une limite : il ne permet pas de cerner l'effet des caractéristiques individuelles, les mesures de fréquence étant basées sur des moyennes d'exposition dans une population, ce qui peut entraîner une interprétation erronée

des résultats que les Anglo-Saxons appellent *ecological fallacy* (erreur écologique). L'erreur écologique est le biais qui survient lorsqu'une association observée entre des variables à l'échelle du groupe ne correspond pas nécessairement à celle qui serait notée à l'échelle individuelle. Une autre limite des études écologiques est leur incapacité de prendre en compte les facteurs de confusion possible.

Les études écologiques sont néanmoins utiles pour décrire des différences dans les populations et permettent de formuler des hypothèses en vue d'études plus poussées. Ainsi, elles peuvent faire ressortir des phénomènes de populations qu'un autre type d'étude descriptive ne saurait cerner. En définitive, il importe que le chercheur accorde une attention particulière à l'adéquation entre ce type d'étude et l'objet de sa recherche.

EXEMPLE 9.13

Une étude a pour objectif d'estimer les taux de mortalité par cancer en fonction du type d'activité industrielle dans des zones géographiques définies. On trouve des taux plus élevés de mortalité par cancer des voies respiratoires dans les zones où des industries pétrolières, du papier et du transport sont en activité. Ces observations mettent les chercheurs sur la piste d'un lien entre ces cancers et certaines des substances potentiellement cancérigènes provenant de ces industries. Si les données de l'étude ne permettent évidemment pas de connaître l'exposition précise à ces substances de chaque individu décédé — pas plus qu'elles ne permettent de tenir compte de l'exposition individuelle à des produits cancérigènes tel le tabac —, elles établissent néanmoins une moyenne basée sur le lieu de résidence par rapport aux industries.

EXEMPLE 9.14

Pendant une épidémie de fièvre dengue, une maladie virale survenant dans les pays tropicaux qui est transmise par le moustique *Aedes aegypti*, on mesure l'incidence de la maladie par quartier de la ville étudiée selon la quantité moyenne de larves observée dans chaque quartier. Dans ce cas, l'unité d'observation qui permet de mieux comprendre l'épidémie n'est pas la quantité de vecteurs de la maladie dans chaque domicile, mais celle observée dans chaque quartier de la ville.

9.3 Les exigences de publication

Danielle, infirmière à la polyclinique, travaille aussi dans un établissement d'enseignement collégial. Elle y rencontre de nombreuses jeunes femmes aux prises avec une infection à *Chlamydia*. Comme elle est désireuse de leur fournir une information utile, elle met la main sur une étude de Massé et autres⁴ intitulée « Infection à *Chlamydia trachomatis* du col de l'utérus : sa prévalence et ses déterminants parmi les femmes qui se présentent pour un examen gynécologique de routine ». Il s'agit d'une étude descriptive réalisée dans un centre local de services communautaires

(CLSC) de Montréal. Danielle décide de faire une lecture critique de l'article afin d'évaluer s'il satisfait ou non aux exigences de publication habituelles. Soucieuse de réaliser cette lecture critique de façon rigoureuse, elle constate qu'une initiative de collaboration internationale impliquant des épidémiologistes, des spécialistes en méthodologie, des statisticiens, des chercheurs et des éditeurs ont uni leurs efforts pour améliorer les publications scientifiques relatives aux études d'observation. Il s'agit de l'initiative internationale STROBE⁵ (*STrengthening the Reporting of OBservational studies in Epidemiology*). Danielle décide de s'inspirer largement de la liste de vérification proposée par ce groupe d'experts internationaux pour faire la lecture critique de l'article.

9.3.1 Le titre et le résumé

L'article propose d'abord un titre et un résumé des différentes sections de l'étude. Un résumé équilibré devrait comprendre les informations relatives à toutes les sections de la publication. Les questions suivantes permettront d'apprécier la pertinence du titre et la qualité de ce type de résumé, appelé « résumé structuré ».

- Le titre et le résumé permettent-ils d'identifier le type ou le modèle d'étude réalisée selon des termes simples et d'usage courant ?
- Retrouve-t-on dans le résumé une information synthétique et objective de ce qui a été fait dans cette étude et de ce qui a été trouvé ?

EXEMPLE 9.15

Danielle, gardant à l'esprit ces questions, prend connaissance de la section « Résumé » de l'article.

Objectif : Déterminer la prévalence et les indicateurs de risque d'infection cervicale à *Chlamydia trachomatis* chez les femmes qui se présentent à un examen médical périodique.

Conception : Étude de prévalence.

Contexte : Nous avons recruté toutes les femmes qui se sont présentées à un examen gynécologique périodique au CLSC Saint-Louis du Parc, à Montréal, de mai 1985 à juillet 1986. Des 773 femmes (99 %) qui ont accepté de participer, 56 ont été exclues, soit parce que les épreuves diagnostiques étaient inadéquates (34), soit parce qu'elles avaient pris des antibiotiques au cours des 6 semaines précédentes (19) ou encore parce que le suivi était impossible après la première visite (3).

Mesures des résultats : La culture était la norme diagnostique, mais on a également utilisé des examens diagnostiques rapides. À partir des cas décelés, on a utilisé des analyses de régression logistique pour évaluer les indicateurs de risque suivants : âge, lieu de résidence, utilisation de contraceptifs oraux, nombre de partenaires sexuels et fréquence des rapports sexuels, antécédents d'infection transmissible sexuellement (ITS) et anomalies constatées à l'examen génital.

EXEMPLE 9.15 (suite)

Principaux résultats : Cinquante et une femmes étaient atteintes d'une infection à *Chlamydia trachomatis*, pour une prévalence de 7,1 % ; 32 femmes (63 %) étaient complètement asymptomatiques. Trois indicateurs indépendants ont été isolés : un âge de 25 ans ou moins (rapport de cote [RC] de 3,2 ; intervalle de confiance [ic] à 95 % de 1,8 à 5,9), un érythème cervical, une hémorragie de contact ou un exsudat muco-purulent (RC de 2,5 ; ic à 95 % de 1,4 à 4,5) et lieu de résidence sur le territoire du CLSC (RC de 2,3 ; ic à 95 % de 1,1 à 5,1). Des antécédents d'ITS ou de vaginite avaient un effet de protection significatif chez les femmes âgées de 30 ans ou plus (RC de 0,2).

Conclusions : Le dépistage des infections à *Chlamydia* pourrait être une mesure de santé publique efficace chez les femmes âgées de 25 ans ou moins et chez celles qui présentent des signes de cervicite au moment d'un test de Papanicolaou.

Danielle considère donc que l'article répond aux critères de qualité en ce qui concerne le résumé structuré. De plus, le contenu de l'article semble répondre à ses besoins professionnels. Elle poursuit donc sa lecture.

9.3.2 L'introduction

L'introduction doit exposer les grandes lignes de l'étude descriptive réalisée et souligner de façon succincte l'intérêt qu'elle revêt en s'appuyant sur des références pertinentes. On y présente, par une revue de la littérature, un bilan de l'état des connaissances en ce qui concerne le problème qui fait l'objet de la recherche. L'auteur définit son approche du sujet ainsi que les objectifs de l'étude.

Les questions suivantes permettront d'apprécier la qualité de l'introduction.

- Le contexte scientifique et la légitimité de l'étude réalisée sont-ils bien indiqués ? En d'autres mots, le raisonnement justifiant la pertinence de l'étude est-il bien conduit ? Ce raisonnement est-il appuyé de façon adéquate par des connaissances scientifiques ?
- Les objectifs spécifiques de l'étude sont-ils clairement énoncés, incluant au besoin les hypothèses *a priori* ?
- Précise-t-on clairement de quel type d'étude descriptive il s'agit ?

EXEMPLE 9.16

Dans l'introduction de l'article, Danielle trouve les informations suivantes appuyées par une dizaine de références.

L'infection par la *Chlamydia trachomatis* est le type d'infection transmissible sexuellement le plus commun en Amérique du Nord. Aux États-Unis, on évaluait à plus de 2,2 milliards de dollars les coûts directs et indirects associés à cette maladie en 1990. Il est difficile de diagnostiquer cette maladie parce que plus

EXEMPLE 9.16 (suite)

de 70 % des femmes infectées sont asymptomatiques. Le dépistage des femmes asymptomatiques pourrait être un moyen de contrôle de la maladie dans la population consultant dans les CLSC.

L'étude transversale réalisée dans ce contexte a donc pour objet de déterminer les indicateurs de risque d'infection du col utérin par la *Chlamydia trachomatis* et sa prévalence chez les femmes consultant ces services de santé de première ligne. On cherche à connaître l'ampleur du problème de santé dans cette population et à cibler les groupes de femmes pour lesquels il pourrait être justifié de procéder à un dépistage systématique.

Danielle considère que la pertinence de l'étude est établie de façon satisfaisante et que ses objectifs sont clairement définis. En outre, les références fournies semblent adéquates et suffisantes pour établir les connaissances scientifiques sur lesquelles s'appuie cette étude et le type d'étude est précisé.

9.3.3 La méthode

La section « Méthode » doit clairement expliquer le choix du type d'étude retenu, décrire la population à l'étude, le procédé au moyen duquel les cas appartenant à la population à l'étude ont été retenus, soit la définition de cas utilisée, la ou les sources de données, les variables retenues, les observations faites et la façon dont celles-ci l'ont été. Cette section devrait aussi contenir un plan d'analyse des données expliquant les biais possibles. On y trouve souvent des précisions sur les méthodes statistiques utilisées. En se fondant sur ces renseignements, le lecteur devrait pouvoir vérifier si la population à l'étude est comparable ou non à celle dont sont issus ses patients.

Danielle retient les éléments suivants de la liste de vérification STROBE pour apprécier la qualité de la section « Méthode ».

- Le modèle d'étude : les éléments clés du modèle d'étude réalisée sont-ils présentés ?
- Le contexte de l'étude : les paramètres essentiels de l'étude sont-ils décrits telles les informations sur le lieu et les dates pertinentes, incluant les informations sur la période de recrutement, d'exposition, de suivi et de collecte des données ?
- Les participants : les critères d'éligibilité, les sources et les méthodes de sélection des participants sont-ils décrits ?
- Les variables : tous les critères de résultats, les expositions, les facteurs de prédictions, les indicateurs, les facteurs confondants potentiels et les facteurs modifiants sont-ils clairement définis ? Les critères retenus pour définir les cas (critères diagnostiques) sont-ils clairement présentés ?
- Les sources de données : pour chaque variable d'intérêt, la source de données et la méthode d'évaluation ou de mesure de la variable sont-elles présentées ? Si des comparaisons entre différents groupes sont réalisées, les méthodes de comparaison sont-elles présentées ?

- Les biais : les mesures prises pour prévenir les sources de biais potentiels sont-elles présentées ?
- La taille de l'échantillon : la méthode ou les critères utilisés pour déterminer la taille de l'échantillon observé sont-ils présentés ? Explique-t-on comment on a déterminé le nombre de sujets à inclure dans l'étude ?
- L'analyse des variables quantitatives : le traitement des variables quantitatives lors de l'analyse est-il expliqué ? Si des regroupements ont été effectués, ces regroupements sont-ils décrits et leur justification est-elle présentée ?
- Les méthodes statistiques : toutes les méthodes d'analyses statistiques utilisées, incluant celle pour contrôler les biais de confusion, sont-elles présentées ? Les méthodes utilisées pour l'analyse des sous-groupes et des interactions sont-elles présentées ? La façon dont on a traité les données manquantes, le cas échéant, est-elle expliquée ? Des **analyses de sensibilité** sont-elles décrites ?

EXEMPLE 9.17

Dans la section « Méthode » de l'article, Danielle peut lire ceci :

On a offert un dépistage de la *Chlamydia trachomatis* à toutes les femmes qui se sont présentées à un examen gynécologique périodique de mai 1985 à juillet 1986 dans un CLSC de Montréal. Des 773 femmes (99 %) qui ont accepté de participer, 56 ont été exclues pour les motifs suivants : les épreuves diagnostiques étaient inadéquates (34) ; des femmes avaient consommé des antibiotiques pendant les 6 semaines précédentes (19) ; le suivi était impossible après la première visite (3). Les analyses sont donc réalisées sur un échantillon de 717 sujets.

Les cas ont été identifiés sur la base du résultat de cultures du col utérin et d'examens diagnostiques rapides. [...] Au cours de l'évaluation de chaque participante, le médecin a rempli un questionnaire standardisé et réalisé un examen physique et gynécologique complet. [...] Un cas a été défini s'il présente au moins une culture positive sur un spécimen recueilli lors de la visite initiale ou lors de la visite de suivi. [...] Des analyses univariées incluant le test du X^2 , le test de Student [...] ont été réalisées. [...] Des analyses de régression logistique* ont été effectuées pour évaluer les indicateurs de risque reconnus dans la littérature, c'est-à-dire l'âge, le lieu de résidence, l'utilisation de contraceptifs oraux, le nombre de partenaires sexuels et la fréquence des rapports sexuels, les antécédents d'ITS et les anomalies constatées à l'examen génital. [...] Des regroupements de variables ont été analysés pour évaluer la sensibilité et la spécificité des modèles prédictifs.

L'essentiel des questions que Danielle doit se poser pour apprécier la qualité de la section « Méthode » trouve une réponse satisfaisante dans l'article. On n'y voit cependant pas d'information explicite sur un calcul d'échantillonnage et les éléments de détail concernant les variables utilisées pour décrire les participantes sont plutôt présentés dans la section résultats.

* Voir dans le chapitre 12 du présent ouvrage la section intitulée « La régression logistique ».

9.3.4 Les résultats

La section « Résultats » doit fournir de l'information sur les caractéristiques générales des individus observés. Des données de fréquence de la maladie en fonction des variables retenues doivent s'y trouver. Selon le cas, les résultats sont présentés sous forme de tableaux et de figures ; les valeurs des tests statistiques y figurent aussi.

Danielle retient les éléments suivants de la liste de vérification STROBE pour apprécier la qualité de la section « Résultats ».

- Les participantes :
 - Le nombre d'individus participants ou non pour les différentes étapes de l'étude est-il présenté? Par exemple, présente-t-on le nombre d'individus susceptibles de participer, le nombre d'individus examinés puis inclus ou exclus, le nombre dont le suivi a été complété et le nombre sur lequel portent les analyses réalisées?
 - Les raisons de la non-participation à certaines étapes de l'étude sont-elles indiquées?
- Les données descriptives :
 - Les caractéristiques des participantes à l'étude (démographiques, cliniques, sociales, etc.) et les informations sur les expositions et les sources de biais de confusion potentielles sont-elles présentées?
 - Le nombre de participantes pour qui l'on observe des données manquantes est-il indiqué pour certaines variables d'intérêt?
 - Les données sur les résultats: le nombre de mesures effectuées et de résultats observés durant la période d'observation est-il présenté?
- Les principaux résultats :
 - Les résultats non ajustés puis ajustés selon les variables de confusion et leur précision (par exemple, l'intervalle de confiance et la valeur p) sont-ils présentés? Clarifie-t-on quel facteur confondant est considéré et pourquoi il a été considéré?
 - Les valeurs bornes des catégories des échelles de classification créées pour les variables quantitatives continues sont-elles indiquées?
- Autres analyses: l'ensemble des analyses réalisées, par exemple les analyses complémentaires par sous-groupes, les analyses d'interaction ou de sensibilité, sont-elles présentées?

EXEMPLE 9.18

À la lecture de la section « Résultats » de l'étude, Danielle constate que le nombre et les caractéristiques des participantes sont clairement présentés dans le tableau 1. Elle retient aussi les éléments suivants :

Cinquante et une femmes étaient atteintes d'une infection à *Chlamydia trachomatis*, pour une prévalence de 7,1 % ; 32 femmes (63 %) étaient complètement

EXEMPLE 9.18 (suite)

asymptomatiques. Trois indicateurs indépendants pouvant constituer des facteurs de risque ont été établis à l'aide de la régression logistique : être âgée de 25 ans ou moins (RC de 3,2; IC à 95 %); présenter une rougeur, une hémorragie de contact ou un exsudat muco-purulent au col utérin (RC de 2,5; IC à 95 %); habiter le territoire du CLSC (RC de 2,3; IC à 95%). Des antécédents d'ITS ou de vaginite avaient un effet de protection significatif pour les femmes âgées de 30 ans ou plus (RC de 0,2).

Comme on le constate, la prévalence de la maladie est précisée. Les indicateurs de risque, présentés sous forme de rapports de cotes (RC), permettent de trouver un antécédent médical ayant un effet protecteur et trois caractéristiques des femmes plus à risque d'être porteuses de l'infection. Les valeurs des tests statistiques et la précision statistique des résultats (intervalles de confiance à 95 %) sont clairement exposées, sauf pour les antécédents d'ITS ou de vaginite. L'analyse par régression logistique utilisée par les auteurs permet de contrôler l'effet potentiellement confondant de certaines variables considérées. Danielle conclut que les réponses satisfaisantes apportées aux questions permettent d'apprécier la qualité de la section « Résultats ».

9.3.5 La discussion

Outre un énoncé des principaux résultats de l'étude, cette section de l'article doit contenir une interprétation de ceux-ci. Une comparaison avec d'autres études permet de mettre en évidence les différences observées à la lumière des connaissances actuelles. Cette section doit présenter les explications et les interprétations possibles ou les limites dans l'interprétation des résultats obtenus, une discussion concernant les validités interne et externe de l'étude et les significations clinique et statistique des résultats. Dans les études de tendance, cette section doit aborder la question des effets de l'âge, de période et de cohorte. On y indique également la portée de l'étude en ce qui a trait aux hypothèses étiologiques ou aux mesures de santé publique qui pourraient en découler. Finalement, c'est dans cette section que les biais sont discutés et que les limites de l'étude sont présentées.

Danielle retient les éléments suivants de la liste de vérification STROBE pour apprécier la qualité de la section « Discussion ».

- Les principaux résultats : les principaux résultats sont-ils résumés en lien avec les objectifs de l'étude?
- Les limites : discute-t-on des limites de l'étude en tenant compte de sources potentielles de biais ou d'imprécision? Discute-t-on de l'effet (surestimation ou sous-estimation de la mesure) et de l'ampleur de cet effet pour chaque source potentielle de biais?
- L'interprétation : discute-t-on d'une interprétation générale prudente des résultats compte tenu des objectifs, des limites de l'étude, de la multiplicité des analyses, des résultats observés dans des études similaires ou d'autres résultats antérieurs pertinents?
- L'inférence statistique : discute-t-on de la validité externe de l'étude, soit de la possibilité d'en généraliser les résultats à une population cible?

EXEMPLE 9.19

Les auteurs de l'étude qu'analyse Danielle indiquent ceci dans la section «Discussion» :

La prévalence des infections à *Chlamydia trachomatis* de 7,1 % confirme l'importance du problème dans la population des femmes s'étant présentées au CLSC pour une consultation. Ce résultat est comparable à ceux obtenus dans les études canadiennes récentes. L'analyse par régression logistique révèle que l'âge est le principal indicateur d'une atteinte (facteur de risque). Ce résultat a également été observé dans plusieurs études antérieures. Cette relation entre l'âge et l'apparition de la maladie n'est pas causale, mais elle traduit probablement un lien avec plusieurs autres facteurs de risque difficilement mesurables, tels l'évolution des comportements sexuels avec l'âge, le recours aux services de santé ou la prédisposition accrue à ce type d'infection.

La présence de biais d'information a été vérifiée en comparant le résultat des examens diagnostiques rapides aux résultats de cultures. L'identification des faux négatifs et des faux positifs a permis d'estimer à 7,6 % la prévalence réelle de la maladie. Les biais de confusion potentiels ont été contrôlés lors de l'analyse à l'aide de la régression logistique. L'étude visait la totalité des femmes qui se sont présentées au CLSC pour une consultation durant la période d'observation. Un biais de sélection pourrait être associé aux femmes qui ont refusé de participer à l'étude. Elles ne représentent cependant que 1 % de celles qui ont demandé une consultation. Le CLSC était situé en milieu urbain, et des études antérieures ont révélé des différences de prévalence entre les milieux rural et urbain. De plus, l'étude a permis de reconnaître des différences de prévalence significatives selon les quartiers étudiés dans la région de Montréal. On doit donc être prudent avant de généraliser les résultats de l'étude à d'autres populations. Le dépistage des infections à *Chlamydia* pourrait être une mesure de santé publique efficace chez les femmes âgées de 25 ans ou moins et chez celles qui présentent des signes de cervicite à l'examen de Papanicolaou. Il serait cependant souhaitable de mener de nouvelles études pour trouver d'autres facteurs de risque et établir la fréquence (périodicité) à laquelle l'examen devrait être effectué.

Danielle trouve dans la section «Discussion» les réponses qui lui permettent d'en apprécier la qualité.

9.3.6 La conclusion

Comme dans beaucoup de publications de ce type, la fin de l'article consulté par Danielle comporte une section conclusion. On devrait y trouver l'essentiel de ce qu'il faut retenir de cette étude, les résultats marquants en lien avec les objectifs de l'étude. D'autres auteurs pourraient présenter ces informations à la fin de la discussion sans toutefois identifier clairement une section «Conclusion». La fin de l'article permettra de répondre aux préoccupations soulevées dans l'introduction. On devrait aussi y trouver des indications sur la direction que pourraient prendre des études futures à la lumière des résultats observés.

Les questions suivantes permettront de répondre à cette préoccupation.

- Quel éclairage nouveau cette étude apporte-t-elle relativement aux préoccupations ayant motivé sa réalisation?
- Les indications sur la direction que pourraient prendre de futures études sont-elles fournies et discutées?

Dans l'étude analysée par Danielle, les auteurs suggèrent ce qui suit, à la lumière des résultats observés : « Le dépistage des infections à *Chlamydia* pourrait être une mesure de santé publique efficace chez les femmes âgées de 25 ans ou moins et chez celles qui présentent des signes de cervicite au moment d'un test de Papanicolaou. »

Danielle retient que cette conclusion des auteurs permet d'orienter la direction de recherches futures portant sur la mesure de l'efficacité du dépistage dans cette population.

Elle considère cependant que l'objectif de l'étude ne portait pas sur la mesure de l'efficacité d'un dépistage éventuel et qu'il est plus prudent d'attendre d'autres études qui en mesureront plus spécifiquement l'efficacité. En définitive, après sa lecture de l'étude de Massé et de ses collègues, Danielle est plus apte à informer adéquatement les jeunes femmes qu'elle rencontre dans cet établissement d'enseignement collégial.

9.3.7 Autre information

Enfin, Danielle constate que les sources de financement ayant permis la réalisation de l'étude sont clairement exposées; les auteurs mentionnent également que le projet d'étude a été soumis à l'approbation d'un comité d'éthique. Cela la rassure quant à la transparence des auteurs. Elle constate cependant que ceux-ci ne donnent pas d'indication concernant les sources potentielles de conflit d'intérêts.

9.4 La surveillance de la santé des populations et de ses déterminants

Après avoir lu le rapport *Surveillance accrue des méningococcies invasives au Canada, 2006 à 2011*⁶, Lisa et Jean se disent heureux de constater que le système de surveillance accrue des méningococcies invasives au pays mis en place par l'Agence de santé publique du Canada leur permet de suivre l'évolution de la fréquence de la maladie dans la population. Ainsi, ils apprennent que le taux d'incidence moyen ajusté pour l'âge pour la période 2006-2011 est de 0,6 nouveau cas pour 100 000 personnes-année pour l'ensemble de la population; il est passé de 0,7 nouveau cas pour 100 000 personnes-année en 2006 à 0,55 nouveau cas pour 100 000 personnes-année en 2011⁷. Ils remarquent cependant que les nourrissons de moins d'un an sont plus touchés avec 7,35 nouveaux cas pour 100 000 personnes-année. Suivent les enfants âgés de 1 à 4 ans avec un taux d'incidence de 1,89 nouveau cas pour 100 000 personnes-année. Ils constatent ainsi l'apport que peuvent avoir de tels systèmes de surveillance sur leurs activités professionnelles.

9.4.1 Une description de la surveillance

S'appuyant sur les modèles et les approches des études descriptives de populations, la surveillance de l'état de santé des populations et de ses déterminants consiste en

plusieurs activités. Il s'agit en fait de recueillir, de produire et de diffuser de façon systématique et régulière une information pertinente afin de dresser un portrait global; d'observer des variations selon le lieu, le temps et les caractéristiques des personnes; de déterminer les problèmes de santé prioritaires; de détecter des problèmes en émergence; d'alerter les autorités quant aux problèmes nécessitant des actions immédiates de protection de la santé; d'élaborer des scénarios prospectifs de l'état de santé; et finalement de suivre l'évolution de certains problèmes précis⁸. L'exercice de la surveillance nécessite aussi la diffusion, en temps opportun, de l'information à ceux qui ont besoin de la connaître.

Les activités de surveillance s'appuient habituellement sur des systèmes informatisés de gestion de l'information qui reposent sur des collectes d'informations spécifiques continues dans les communautés (surveillance à base communautaire) ou dans le système de santé (hospitalisations), périodiques (enquêtes) ou encore sur des bases de données administratives existantes. Ces différentes sources de données pourront également s'avérer complémentaires.

Un système de surveillance se distingue d'un outil de collecte d'information périodique dans la mesure où, par la mise à jour systématique des données selon une périodicité établie, un tel système permet la normalisation des bases de données. Cette caractéristique assure des comparaisons valides des résultats selon le lieu, le temps et les caractéristiques des personnes, de même qu'une normalisation des méthodes statistiques et épidémiologiques utilisées pour produire ces mesures de façon systématique. Le recensement canadien réalisé tous les cinq ans est un exemple connu de tous les citoyens qui répond normalement à ces critères de qualité et qui permet d'utiliser les données recueillies à des fins de surveillance.

Un système de surveillance peut s'appuyer tantôt sur des sources de données couvrant la population totale du territoire ciblé, tantôt sur des processus de «surveillance sentinelle» en comptant plutôt, par exemple, sur la participation de polycliniques, de pharmacies ou de centres hospitaliers sélectionnés.

EXEMPLE 9.20

On veut connaître l'évolution de la fréquence et de la répartition du cancer du sein à l'aide des données du système d'information qui permet de soutenir les activités quotidiennes d'un programme de dépistage. Ce système permet l'enregistrement, le traitement et la circulation de l'information relative au questionnaire médical des participantes, à la réalisation de la mammographie de dépistage et à la transmission des résultats. Bien qu'il puisse constituer une source d'information importante, ce système à lui seul ne fournira pas une mesure exhaustive du cancer du sein dans la population parce qu'il ne génère que les données relatives aux participantes au programme de dépistage. Pour obtenir un portrait exhaustif de la fréquence et de la répartition du cancer du sein dans la population, les données du système d'information du dépistage devront être complétées par d'autres données (par exemple, celles du *Registre québécois du cancer* ou d'une base de données administrative du système de santé qui regroupent les renseignements portant sur les diagnostics et les hospitalisations). Ces trois sources de données gagneront

EXEMPLE 9.20 (suite)

également à être étoffées par une collecte d'information complémentaire essentielle, mais absente de ces sources, sur le grade de cancer, par exemple. Lisa et Jean constatent ainsi que les mesures épidémiologiques issues des activités et des systèmes de surveillance sont d'un intérêt et d'une utilité comparables à celles des études descriptives. En fait, ces systèmes de surveillance peuvent constituer des sources d'information utile dans le cadre d'études descriptives de populations, et ce, dans divers domaines de la santé.

9.4.2 Les objets de surveillance

Les objets de surveillance, ou indicateurs de santé et de bien-être, sont déterminés en tenant compte des enjeux de santé et de bien-être dans la population ciblée. Ces objets de surveillance peuvent couvrir l'ensemble des dimensions de la santé des populations et de ses déterminants. Le tableau 9.1 présente les grandes classes d'objets de surveillance.

TABLEAU 9.1 ■ Un exemple de classes d'objets de surveillance

Déterminants de la santé	Incapacité
Conditions démographiques	Morbidité
Conditions socioéconomiques	Mortalité
Conditions socioculturelles	Limitations fonctionnelles
Environnement social	État de santé psychosociale/santé mentale
Environnement physique	Problèmes de santé physique
Habitudes de vie et comportements	Problèmes liés à la santé des nouveau-nés et des nourrissons
Caractéristiques biologiques et génétiques	Traumatismes
Systemes et services de santé	Maladies infectieuses
	Intoxications d'origine environnementale
	Affections liées au milieu de travail

9.4.3 L'utilité de la surveillance

Comme pour les études descriptives de populations, les mesures épidémiologiques réalisées dans le cadre des activités de surveillance de la santé des populations et de ses déterminants visent à informer la population sur son état de santé et ses déterminants et ainsi éclairer ses choix en matière de santé. Ces mesures permettent également de soutenir les intervenants et les décideurs dans leur processus de planification et d'élaboration de politiques, de programmes ou de plans d'action.

Par la production et la diffusion de mesures épidémiologiques en continu — ou mises à jour périodiquement —, la surveillance de l'état de santé des populations permet aussi de soutenir les cliniciens dans le processus diagnostique. D'une part, elle les informe de la fréquence et de la répartition des maladies dans la population selon le

lieu, le temps et les caractéristiques des personnes. D'autre part, elle peut contribuer à déterminer lesquelles des pratiques cliniques préventives sont les plus pertinentes compte tenu des caractéristiques et des besoins des populations. Ainsi, on dira que la surveillance de la santé constitue «un élément essentiel des pratiques en matière de décision fondée sur des preuves⁹».

9.4.4 Des exemples de systèmes de surveillance

La surveillance de l'état de santé au Québec

Le Québec produit, à intervalles de cinq ans, une grande étude descriptive de la santé de sa population et de ses déterminants. Une première édition a été produite en 2001, une deuxième en 2006 et une troisième en 2011¹⁰. Cette dernière étude présentait 200 mesures épidémiologiques et statistiques de la santé des Québécois et de ses déterminants réalisées à partir de 70 bases de données différentes. Elle a permis de valider, d'une part, que l'espérance de vie est en hausse constante au Québec, se situant parmi les plus élevées au monde, et, d'autre part, que l'espérance de vie en bonne santé des Québécois est l'une des plus élevées au Canada. Cette étude traduit également un recul important de la mortalité infantile, une diminution importante de la mortalité par maladie de l'appareil circulatoire et une réduction constante du tabagisme. De plus, on y observe que le cancer constitue désormais la première cause de décès et que la mortalité par cancer du poumon a doublé chez les femmes. Elle fait état d'une progression importante de l'obésité, révèle une augmentation des cas rapportés de négligence chez les enfants, une mortalité par suicide très élevée et une disparité importante de la santé selon les régions, le statut socioéconomique, le sexe et l'appartenance à une nation autochtone.

En complément au document papier, ces 200 indicateurs ont été repris dans le cadre d'un système de surveillance. Mis à jour selon une périodicité adaptée aux différents indicateurs, et accessibles par les professionnels de santé publique autorisés sur un portail informationnel, ils peuvent être exploités de façon dynamique en ligne à l'aide de requêtes simples s'appuyant sur des progiciels statistiques et cartographiques; ces derniers permettent de produire des résultats adaptés aux besoins des chercheurs, des planificateurs et des décideurs. Un processus de mise à jour systématique des données selon une périodicité établie, une normalisation des bases de données permettant des comparaisons valides des résultats selon le lieu, le temps et les caractéristiques des personnes, et une normalisation des méthodes statistiques et épidémiologiques sont autant de caractéristiques qui permettent de parler de système de surveillance de l'état de santé de la population et de ses déterminants. S'appuyant sur ce système de surveillance, le Directeur national de la santé publique du Québec publiait en 2016 un quatrième rapport sur l'état de santé des Québécois¹¹.

Santé Canada et l'Organisation mondiale de la Santé

Sur le site Web de l'Agence de santé publique du Canada, on peut accéder à 25 programmes et systèmes de surveillance couvrant des domaines de santé précis et répondant à des besoins particuliers de connaissance, de contrôle et d'intervention pour la planification et la protection de la santé des populations.

Les systèmes de surveillance peuvent avoir une portée régionale, nationale ou internationale, comme le système de surveillance de la grippe à l'Organisation mondiale de la Santé (OMS).

L'OMS œuvre à la surveillance et au contrôle de la grippe depuis 1948¹². Cette surveillance a évolué jusqu'à permettre la mise en place d'un système de surveillance de cette maladie infectieuse dans le cadre d'un partenariat mondial. Depuis 1998, ce système incite 110 laboratoires, dans 82 pays, à surveiller continuellement les virus grippaux isolés localement et à s'informer mutuellement de l'émergence et de la propagation des différentes souches. Grâce à l'information ainsi recueillie, les experts associés au système de surveillance peuvent formuler des recommandations sur les souches virales à inclure dans le vaccin antigrippal de la saison suivante. Ainsi, les données issues de la surveillance mondiale soutiennent les actions de santé publique et l'élaboration des plans nationaux et mondiaux de lutte systématique. Elles soutiennent également les interventions cliniques en permettant la fabrication de vaccins plus efficaces et en favorisant la mise en place d'actions rapides et ordonnées à l'échelle mondiale dans les cas d'émergence d'épidémies ou de **pandémies**.

9.4.5 Les enjeux éthiques des études descriptives de populations et de la surveillance

La réalisation des études descriptives et des activités de surveillance ainsi que la mise en place de systèmes de surveillance soulèvent des enjeux éthiques incitant à une grande vigilance. On doit notamment assurer la sécurité des données et la protection des renseignements personnels. Pour ce type d'étude, on doit parfois procéder à la collecte des données avec le consentement éclairé et la participation volontaire des sujets. L'accès aux données existantes nécessite transparence et respect des règles et des lois relatives à la sécurité des données et à la protection des renseignements personnels.

La réalisation d'études descriptives de populations et la mise en place de systèmes de surveillance s'appuyant sur des bases de données existantes peuvent requérir le jumelage de bases de données, accentuant ainsi le risque de non-respect de la vie privée et de la protection des renseignements personnels.

Enfin, on doit procéder à l'exploitation des données et à la diffusion des résultats issus de ces activités en veillant à ne pas nuire aux individus observés, c'est-à-dire en évitant qu'ils soient indirectement identifiables par des résultats présentant des petits nombres, à l'aide de caractéristiques individuelles, de lieux et de temps. On doit aussi veiller à éviter de stigmatiser des sous-groupes de populations en utilisant notamment un vocabulaire approprié pour les risques liés à certaines de leurs caractéristiques telles l'appartenance religieuse ou les pratiques sexuelles.

En résumé

L'étude descriptive de populations s'attarde à une population bien déterminée qu'elle analyse en fonction des variables retenues. On distingue deux principaux types d'études descriptives de populations : les études transversales (ou études de prévalence) et les études longitudinales (études d'incidence, de mortalité, de létalité ou de survie). Si le chercheur observe une population à divers intervalles, il mène une « étude de tendance ». L'étude écologique constitue également une étude descriptive de populations dont la particularité est de tenir compte d'un groupe d'individus plutôt que de considérer l'individu isolément.

L'étude descriptive de populations sert d'abord à connaître la fréquence et la répartition de la maladie ou du problème d'intérêt dans la population selon les caractéristiques de personnes, de temps et de lieux. S'appuyant sur les modèles et les approches des études descriptives de populations, la surveillance de l'état de santé des

populations et de ses déterminants permet les réalisations suivantes : dresser de façon périodique ou en continu un portrait global de la santé des populations et de ses déterminants ; observer des variations et des tendances selon le lieu, le temps et les caractéristiques des personnes ; identifier les problèmes de santé prioritaires ; détecter des problèmes en émergence ; alerter les autorités quant aux problèmes nécessitant des actions immédiates de protection de la santé ; élaborer des scénarios prospectifs de l'état de santé ; et suivre l'évolution de certains problèmes de santé.

En santé publique, l'étude descriptive de populations ainsi que la surveillance de la santé de la population et de ses déterminants permettent de cibler des populations à risque alors qu'en recherche épidémiologique, elles permettent de dégager les connaissances de base nécessaires à la formulation d'hypothèses étiologiques.

Exercices



Voir la section complète des exercices sur le Web.

- Quel type d'étude descriptive devrait-on utiliser pour mesurer la fréquence des nouveaux cas de cancer et les caractéristiques des personnes atteintes ?
 - Une étude descriptive de population transversale.
 - Une étude de tendance.
 - Une étude descriptive de population longitudinale
 - Une étude écologique.
- Vrai ou faux. Dans une étude descriptive de population, le groupe témoin est issu de la population à l'étude.
- Dans une publication qui présente les résultats d'une étude descriptive de populations, dans quelle section devrait-on présenter les limites des résultats observés ?
 - Dans la section « Introduction ».
 - Dans la section « Méthode ».
 - Dans la section « Résultats ».
 - Dans la section « Discussion ».
 - Dans la section « Conclusion ».
- Complétez la phrase suivante.
La surveillance et les études descriptives de population servent d'abord à connaître la _____ et la _____ de la maladie, des décès et de leurs déterminants selon des caractéristiques de _____, de _____ et de _____.

Pour les réponses, voir le corrigé à la fin du manuel.

Collaboration à la 3^e édition :
Suzanne Gingras

Au terme de ce chapitre, vous pourrez :

- expliquer les concepts de normalité statistique et d'inférence statistique ;
- prendre une décision éclairée, en vous appuyant sur les principes de l'inférence statistique, quant à la pertinence de modifier votre pratique clinique à partir de résultats présentés dans des études scientifiques.

Objectifs spécifiques	Contenu
10.1 Définir une probabilité.	10.1 La notion de probabilité 10.1.1 La probabilité objective: fréquentiste, <i>a priori</i> 10.1.2 Les propriétés d'une probabilité objective
10.2 Distinguer une population d'un échantillon.	10.2 Les notions de population et d'échantillon 10.2.1 La population 10.2.2 L'échantillon
10.3 Expliquer l'inférence statistique: distribution de probabilités.	10.3 L'inférence 10.3.1 Le test d'hypothèse 10.3.2 L'intervalle de confiance 10.3.3 La distribution de probabilités
10.4 Distinguer la normalité statistique de la normalité clinique.	10.4 La normalité statistique et la normalité clinique 10.4.1 La normalité statistique 10.4.2 La normalité clinique





Mise en situation

Charles consulte Lisa, omnipraticienne à la polyclinique Milo, afin qu'elle prenne sa pression artérielle, qu'il croit anormale. En effet, quelques jours plus tôt, à l'occasion d'une campagne de dépistage de l'hypertension artérielle (HTA) dans son milieu de travail, on a mesuré sa pression. La mesure indiquait 150/98 mmHg; on lui a donc recommandé de consulter son médecin de famille. Charles a 48 ans, se croit en très bonne santé et ne présente aucun symptôme. En revanche, il fume un paquet de cigarettes par jour depuis l'âge de 20 ans. Sa mère, qui souffrait d'HTA, est décédée des suites d'un accident vasculaire cérébral. La taille et le poids de Charles se situent dans les limites de la normale. Lisa mesure sa pression artérielle à trois reprises selon les méthodes recommandées; elle est élevée, se situant à 156/100 mm Hg. Lisa lui prescrit quelques analyses de laboratoire visant à compléter l'investigation. Elle lui demande de faire mesurer sa pression artérielle au moins une fois par semaine d'ici le prochain rendez-vous, dans six semaines. Elle lui recommande également de réduire sa consommation de cigarettes et de sel, de bien dormir et de se détendre.

Pertinence clinique

Un cas comme celui de Charles conduit Lisa à s'interroger. L'anomalie de la pression artérielle que présente Charles se rencontre-t-elle fréquemment? Comporte-t-elle des risques pour la santé? Quelle est la probabilité qu'un individu présente les mêmes valeurs de pression artérielle que Charles? Pour répondre à ces questions, Lisa peut certes se fier à son expérience clinique et à ses connaissances. Elle peut aussi consulter la littérature scientifique pour mieux reconnaître les problèmes liés à l'HTA. Ce faisant, Lisa est mise en présence d'études qui décrivent la répartition des différentes valeurs de la pression artérielle dans la population ou qui tentent d'établir des relations entre certaines valeurs de la pression artérielle et la survenue de problèmes de santé. Pour bien utiliser l'information transmise dans ces études, Lisa doit posséder les connaissances indispensables à la compréhension des études statistiques en général. Ces notions essentielles font l'objet du présent chapitre et des deux chapitres suivants.

10.1 La notion de probabilité

Une **probabilité** peut être définie de façon intuitive; il s'agit alors d'une estimation personnelle ou d'une impression basée sur l'expérience. Une probabilité peut aussi être définie de façon objective: on parle alors de «probabilité fréquentiste» et de «probabilité *a priori*».

10.1.1 La probabilité objective : fréquentiste, *a priori*

Dans l'approche « fréquentiste » des probabilités, une expérience est répétée un très grand nombre de fois et, chaque fois, un ou plusieurs événements se produisent. Chacune des expériences constitue un événement aléatoire puisque l'on ne peut en prévoir le résultat avec certitude, bien que l'on puisse en décrire tous les résultats possibles. La probabilité qu'un certain événement survienne s'exprime comme suit.

$$\frac{\text{Nombre de fois que l'événement s'est produit}}{\text{Nombre total d'événements}}$$

EXEMPLE 10.1

Supposons qu'on lance un dé 100 fois et que l'on obtienne les résultats suivants :

Face du dé	Nombre de lancers	Face du dé	Nombre de lancers
1	20	4	17
2	16	5	16
3	13	6	18

On remarque que la proportion d'obtention du 1 est de 20/100, alors que celle du 5 est de 16/100. On peut donc dire que les probabilités d'obtenir le 1 et le 5 sont respectivement estimées à 20/100 et 16/100.

Une probabilité peut aussi être définie de façon objective *a priori*. S'il s'agit d'événements dont la probabilité de survenir est égale, elle s'exprime alors de la manière suivante.

$$\frac{\text{Nombre de fois qu'un événement peut se produire}}{\text{Nombre total d'événements possibles}}$$

EXEMPLE 10.2

- Six événements sont possibles lorsqu'un dé est lancé : obtenir le 1, le 2, le 3, le 4, le 5 ou le 6. À chaque lancer, un seul événement peut se produire ; la probabilité d'obtenir le 1 est donc de 1/6. La probabilité d'obtenir le 5 est également de 1/6. Cela n'est évidemment vrai que si le dé est bien équilibré.
- Lorsqu'une carte est choisie au hasard dans un jeu de cartes, quelle est la probabilité d'obtenir le 8 de pique ? Comme il y a 52 cartes dans un jeu, il y a 52 événements possibles. La probabilité d'obtenir le 8 de pique au cours d'un essai est donc de 1/52. Quelle est la probabilité d'obtenir le 5, peu importe la couleur ? Puisque 4 cartes portent le chiffre 5 (pique, cœur, carreau et trèfle), cette probabilité est de 4/52, soit de 1/13.
- Une future mère aux yeux bleus et un futur père aux yeux bruns s'interrogent sur ce que sera la couleur des yeux de leur enfant. La génétique indique que chacun des parents fournit un gène à l'enfant, que le gène brun (que l'on désignera par

EXEMPLE 10.2 (suite)

la lettre «B») est dominant et que le gène bleu (que l'on désignera par la lettre «b») est récessif. Par conséquent, la mère a le génotype bb puisqu'elle a les yeux bleus; elle transmettra le gène b à son enfant. Le père a le génotype Bb ou le génotype BB. S'il a le génotype Bb, il transmettra soit le gène b, soit le gène B à son enfant. Les événements possibles sont donc: bb (yeux bleus) ou Bb (yeux bruns). La probabilité que l'enfant ait les yeux bleus (bb) est donc de 1/2, soit de 50 %, si le père est porteur du génotype Bb. Elle est évidemment de 0 % s'il est porteur du génotype BB.

10.1.2 Les propriétés d'une probabilité objective

Une probabilité objective comporte des propriétés, dont les trois suivantes.

Premièrement, la probabilité de survenue de chaque événement varie entre 0 et 1.

EXEMPLE 10.3

La probabilité que survienne un événement impossible, comme celle de lancer un dé et d'obtenir le 7, est de 0, soit de 0/6, car le chiffre 7 n'est représenté sur aucune face du dé. La probabilité que survienne un événement certain, comme celle d'obtenir une valeur de 1 à 6 au lancer du dé, est de 1, soit de 6/6, car cet événement se produira nécessairement à chaque essai. La probabilité de tous les autres événements ou groupes d'événements possibles se situe entre ces deux extrêmes.

Deuxièmement, la somme des probabilités de tous les événements possibles pour une expérience est 1.

EXEMPLE 10.4

Reprenons l'expérience présentée dans l'exemple 10.1. Tous les événements possibles au cours du lancer d'un dé sont d'obtenir le 1, le 2, le 3, le 4, le 5 ou le 6. La probabilité associée à chacun des événements de cette expérience est respectivement de 20/100, 16/100, 13/100, 17/100, 16/100 et 18/100. La somme des probabilités de tous les événements possibles pour cette expérience est donc de:

$$\frac{20+16+13+17+16+18}{100} = \frac{100}{100}, \text{ soit } 1.$$

Troisièmement, chaque expérience est dite «dépendante» si elle influence le résultat des expériences subséquentes. Au contraire, elle est dite «indépendante» si le résultat obtenu n'a pas d'influence sur celui des expériences subséquentes.

EXEMPLE 10.5

Un 5 est choisi au hasard dans un jeu de cartes. La probabilité d'obtenir un 5 était de 4/52, soit de 1/13. Si la carte n'est pas replacée dans le jeu, la probabilité d'obtenir un autre 5 est de 3/51 pour l'expérience subséquente, car le jeu ne compte plus que 51 cartes, dont 3 seulement portent le chiffre 5. Ces deux expériences sont donc dépendantes.

EXEMPLE 10.5 (suite)

En revanche, si la carte est replacée dans le jeu, la probabilité d'obtenir un 5 est à nouveau de $1/13$; le résultat de cette expérience n'est pas influencé par le résultat précédent. Ces expériences sont donc indépendantes.

La notion de probabilité comporte plusieurs autres lois qui ne sont pas présentées ici. Elles concernent notamment l'addition et la multiplication des probabilités selon qu'il y a indépendance ou dépendance entre les événements¹.

À retenir

- Les probabilités «fréquentiste» et «*a priori*» sont des probabilités objectives.
- La probabilité de survenue d'un événement varie entre 0 et 1. La somme des probabilités de tous les événements possibles pour une expérience est de 1.
- Une expérience est dite «indépendante» si le résultat obtenu n'influence pas celui des autres expériences.

10.2 Les notions de population et d'échantillon**10.2.1 La population**

En statistique, le terme **population** est utilisé pour désigner un ensemble d'éléments (des individus ou des objets) dotés de caractéristiques communes.

EXEMPLE 10.6

- Une étude porte sur tous les patients de la polyclinique Milo. Chacun des patients est un élément et l'ensemble des patients constitue la population.
- Danielle, l'infirmière à la polyclinique, s'intéresse au nombre d'enfants vaccinés contre la rougeole. La population qu'elle étudie est formée des élèves d'une école où elle travaille également. Les éléments sont les enfants participant à l'étude.
- Une population peut être constituée de tous les dossiers de la polyclinique, des lames du laboratoire de cytologie d'un hôpital ou des différentes pièces anatomiques qui sont apportées au laboratoire de pathologie d'un hôpital. Les éléments sont alors respectivement les dossiers, les lames et les pièces anatomiques.
- Une étude porte sur la proportion d'ordonnances de narcotiques par rapport aux autres médicaments prescrits au cours d'une même journée. La population est formée de toutes les ordonnances remplies dans une pharmacie pendant cette journée. Chacune des ordonnances constitue l'un des éléments.

La notion de paramètre

Lorsqu'on mesure la moyenne d'une variable observée chez tous les individus d'une population, on obtient le **paramètre**, c'est-à-dire la moyenne de cette

variable dans la population en question. Il en est de même pour toute autre mesure de tendance centrale ou de dispersion telle que la médiane, le mode ou l'écart type.

EXEMPLE 10.7

Lisa décide de mesurer la pression artérielle (la variable observée) de tous les hommes qui se présentent à la polyclinique Milo le lendemain de la visite de Charles. Cet ensemble d'individus constitue la population à l'étude. Lisa obtient les résultats présentés dans le tableau 10.1.

Pour les 10 hommes âgés de 37 ans à 71 ans, la moyenne de la pression artérielle diastolique est de 78 mm Hg et l'écart type est de 2,4 mm Hg. La moyenne de la pression artérielle systolique est de 125 mm Hg et l'écart type est de 6 mm Hg. Ces mesures sont les paramètres de cette population.

TABLEAU 10.1 ■ La pression artérielle de la population à l'étude

Numéro	Âge (ans)	Pression artérielle systolique (mm Hg)	Pression artérielle diastolique (mm Hg)	Numéro	Âge (ans)	Pression artérielle systolique (mm Hg)	Pression artérielle diastolique (mm Hg)
1	37	115,0	74,0	6	52	126,0	78,0
2	41	118,0	75,0	7	56	128,0	79,0
3	46	120,0	76,0	8	61	130,0	80,0
4	47	122,0	77,0	9	65	132,0	81,0
5	49	124,0	78,0	10	71	135,0	82,0

10.2.2 L'échantillon

Lorsqu'on s'intéresse à un sous-ensemble d'une population, on constitue un échantillon.

EXEMPLE 10.8

Prenons le sous-ensemble que présente le tableau 10.2, qui a été formé à partir de la population dont il est question dans le tableau 10.1.

TABLEAU 10.2 ■ Un échantillon constitué à partir de la population à l'étude

Numéro	Âge (ans)	Pression artérielle systolique (mm Hg)	Pression artérielle diastolique (mm Hg)
2	41	118,0	75,0
7	56	128,0	79,0

Cet ensemble d'individus constitue un échantillon de la population à l'étude.

Un échantillon peut être obtenu à partir de n'importe quelle population. Ainsi, certains dossiers de la polyclinique, une classe d'élèves d'une école ou une série de lames provenant du laboratoire de cytologie peuvent constituer des échantillons.

Afin d'étudier un problème, il est le plus souvent indispensable de constituer un échantillon à partir de la population à l'étude. En effet, pour des raisons liées aux coûts ou aux contraintes de temps, il est souvent impossible d'étudier l'ensemble des éléments d'une population. L'étude d'un échantillon permet de concentrer les efforts pour obtenir les meilleurs résultats possible. En outre, on peut utiliser dans l'étude d'un échantillon des instruments de mesure de meilleure qualité, en particulier si les ressources humaines et financières sont limitées. De ce fait, les résultats obtenus à partir de l'étude de l'échantillon s'avèrent souvent plus adéquats que ceux que l'on obtiendrait en étudiant toute la population.

EXEMPLE 10.9

Un chercheur s'intéresse à la relation entre la pression artérielle et l'apparition de maladies cardiovasculaires. Or, il existe un appareil ambulatoire d'enregistrement de la pression artérielle; mais cet appareil est coûteux, et il doit être porté jour et nuit par la personne chez qui l'on effectue la mesure. Par conséquent, on ne peut l'utiliser pour étudier l'ensemble des individus d'une population. Cependant, en constituant un échantillon à partir de cette population, on rend possible l'utilisation de l'appareil tout en limitant les frais et l'on obtient une mesure dont la qualité est grandement augmentée.

Il existe deux catégories de population: la **population à l'étude**, soit celle à partir de laquelle le chercheur sélectionne des éléments pour former un échantillon, et la **population cible**, qui est celle à laquelle il veut généraliser ses résultats. En milieu clinique, la population à l'étude est souvent constituée des patients plus facilement accessibles et qui répondent à certains critères d'admissibilité.

EXEMPLE 10.10

Dans l'étude que mène Lisa, la population cible est constituée des hommes qui se présentent à la polyclinique Milo. La population à l'étude se compose des hommes qui se présentent à la polyclinique Milo le lendemain de la visite de Charles. L'échantillon est formé de deux individus faisant partie de cette population.

Les résultats obtenus à partir de la population à l'étude s'appliquent-ils à la population cible? Oui, à la condition que la population à l'étude soit comparable à la population cible, ce que l'on peut vérifier en s'assurant que la distribution de certaines variables importantes est semblable dans les deux populations. Ce jugement est clinique et non statistique; il confirme la validité externe de l'étude.

EXEMPLE 10.11

Il est important de s'assurer que l'âge des patients qui se présentent à la polyclinique Milo le lendemain de la visite de Charles est comparable à l'âge de ceux qui y viennent de manière habituelle.

EXEMPLE 10.11 (suite)

De plus, comme la population à l'étude est formée uniquement d'hommes, les résultats que Lisa obtiendra ne pourront être appliqués qu'à une population cible masculine, soit les hommes qui se présentent à la polyclinique Milo. De même, lorsque Lisa consultera la littérature scientifique, elle devra définir la population cible des études publiées, de manière à savoir à qui les résultats obtenus peuvent s'appliquer. Ainsi, elle ne pourra pas utiliser les résultats d'une étude décrivant les différentes valeurs de la pression artérielle obtenues dans un échantillon d'hommes âgés de 20 ans à 24 ans pour évaluer la probabilité d'obtenir, en ce qui concerne Charles, une valeur de pression artérielle de 156/100 mm Hg. Elle devra plutôt consulter une étude effectuée à partir d'un échantillon d'hommes âgés de 45 ans à 49 ans, un groupe correspondant à celui auquel appartient Charles ; en effet, la pression artérielle augmente avec l'âge.

L'**échantillon électif** est formé d'individus sélectionnés par le chercheur de façon non aléatoire. C'est souvent le cas des échantillons utilisés pour des études à visée étiologique. Dans de telles recherches, les échantillons sont formés en tenant compte de critères de similitude entre les individus pour un certain nombre de variables, sauf pour la variable qui fait l'objet de l'étude.

EXEMPLE 10.12

Une étude porte sur la relation entre les maladies cardiovasculaires et la pression artérielle. Les individus sélectionnés sont comparables en ce qui concerne l'âge, le sexe, la consommation de tabac et les habitudes alimentaires, mais les valeurs de la pression artérielle observées chez eux diffèrent. Si une différence dans l'apparition de la maladie cardiovasculaire est notée, elle sera plus certainement associée au niveau de la pression artérielle qu'aux autres variables.

Certains résultats provenant de l'étude d'un échantillon électif peuvent ne pas s'appliquer à l'ensemble de la population à l'étude. Les caractéristiques des individus de l'échantillon ne sont pas forcément comparables à celles des individus de l'ensemble de la population à l'étude, car l'échantillon a été choisi plutôt que formé au hasard. Ainsi, dans l'exemple précédent, il serait possible de compter une proportion de 75 % de fumeurs dans les deux groupes, ce qui n'est certainement pas représentatif de ce que l'on peut observer dans la population. Par contre, les associations observées, c'est-à-dire, dans ce cas, le lien (risque relatif ou rapport de cotes) entre les maladies cardiovasculaires et la pression artérielle, sont applicables à la population à l'étude.

La notion de statistique

Une **statistique** est une valeur mesurée dans un échantillon. Cette valeur peut être la moyenne, l'écart type, la médiane, etc. Des statistiques différentes sont calculées pour chaque échantillon constitué à partir d'une population.

EXEMPLE 10.13

Dans l'échantillon formé des individus 2 et 7 présenté dans le tableau 10.2, la moyenne échantillonnale, c'est-à-dire la statistique, de la pression artérielle diastolique est de 77 mm Hg. Si l'échantillon est formé des individus 1 et 3, la moyenne devient 75 mm Hg.

Les principales différences entre un paramètre (dans ce cas, celui mesuré dans la population à l'étude, soit tous les hommes qui se présentent à la polyclinique Milo le lendemain de la visite de Charles) et une statistique (dans ce cas, celle mesurée dans l'échantillon formé à partir de cette population) sont présentées dans le tableau 10.3.

TABLEAU 10.3 ■ Les principales différences entre le paramètre mesuré et la statistique mesurée

	Paramètre	Statistique
Source	Population	Échantillon (individus 2 et 7)
Moyenne		
Notation	μ	\bar{x}
Valeur	78,0 mm Hg	77,0 mm Hg
Écart type		
Notation	σ	s
Valeur	2,4 mm Hg	2,0 mm Hg

À retenir

- La population est l'ensemble des éléments dotés de caractéristiques communes.
- L'échantillon est un sous-ensemble de la population. S'il est choisi au hasard, on le dit aléatoire ; il peut alors être représentatif de la population. S'il est choisi autrement, on le dit électif ; il risque alors de ne pas être représentatif de la population.
- Un paramètre est une mesure effectuée sur la population.
- Une statistique est une mesure effectuée sur un échantillon tiré de la population.

10.3 L'inférence

Comme l'étude de l'ensemble de la population est souvent impossible, le paramètre demeure inconnu. L'estimation de ce paramètre consiste à supposer que sa valeur s'apparente à celle de la statistique correspondante mesurée dans l'échantillon. Dans l'exemple 10.13, on a comme estimation de la moyenne de la population 77 mm Hg avec l'échantillon formé des individus 2 et 7, et 75 mm Hg avec celui formé des individus 1 et 3. La valeur de la statistique représente l'estimation ponctuelle du paramètre. Cependant la probabilité que le paramètre soit égal à la statistique s'avère très faible. À partir d'un échantillon, on désire souvent estimer un paramètre d'une population ou tester une hypothèse de recherche, alors que peut-on faire si la probabilité que la valeur d'une statistique soit égale au paramètre correspondant est faible ?

L'**inférence** est le processus par lequel on induit la valeur d'un paramètre à partir de la statistique correspondante mesurée dans un échantillon représentatif de la

population. Deux méthodes d'inférence sont basées sur la théorie des probabilités : le test d'hypothèse et l'intervalle de confiance.

10.3.1 Le test d'hypothèse

L'approche du **test d'hypothèse** consiste à soupeser deux hypothèses faites à l'échelle de la population à partir des données d'un échantillon. Les deux hypothèses soupesées sont liées aux objectifs de l'étude et peuvent prendre différentes formes. Par exemple, pour vérifier si la moyenne de la pression artérielle chez les femmes est la même que celle chez les hommes, on doit confronter l'hypothèse que les deux moyennes sont égales avec celle que les deux moyennes sont différentes. Si l'on voulait plutôt vérifier si la moyenne chez les femmes est inférieure à celle chez les hommes, on devrait confronter cette dernière hypothèse avec l'hypothèse que la moyenne chez les femmes est égale ou supérieure à celle chez les hommes. Le test d'hypothèse est une démarche qui permet de retenir ou non une des deux hypothèses, avec une certaine marge d'erreur. La notion de test d'hypothèse sera définie et expliquée plus en détail dans le chapitre 11 (*voir la section intitulée « La réalisation d'un test d'hypothèse »*).

10.3.2 L'intervalle de confiance

Un intervalle peut aussi être construit autour de l'estimation ponctuelle. C'est ce que l'on appelle l'**intervalle de confiance**. Celui-ci permet de conclure, avec une certaine marge d'erreur, quant à la valeur du paramètre dans la population à partir de l'analyse de l'échantillon. De plus, dans certaines situations, l'intervalle de confiance peut être utilisé pour faire un choix entre deux hypothèses, donc pour effectuer un test d'hypothèse. La notion d'intervalle de confiance sera définie et expliquée plus en détail dans le chapitre 11 (*voir la section intitulée « L'intervalle de confiance »*).

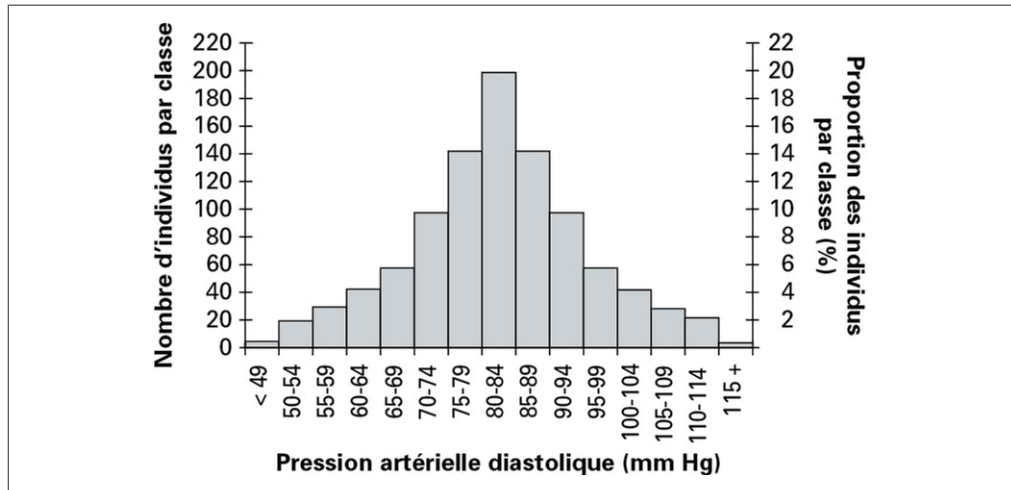
10.3.3 La distribution de probabilités

Afin d'estimer correctement un paramètre, sa **distribution de probabilités** doit être connue. Or, qu'est-ce qu'une distribution de probabilités ? C'est un modèle de représentation de la fréquence des différentes valeurs d'une variable dans un échantillon ou une population.

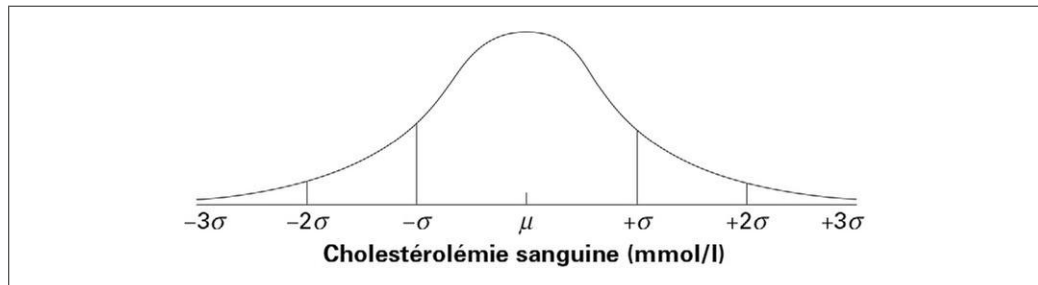
Certaines distributions de probabilités sont importantes en statistique ; c'est le cas, plus particulièrement, de la distribution dite « normale ».

EXEMPLE 10.14

Dans la ville de Québec, les valeurs de la pression artérielle diastolique sont mesurées chez 1 000 individus. Pour construire la distribution de probabilités, les différentes valeurs de la pression artérielle sont regroupées en classes. Le nombre d'individus que compte chaque classe est ensuite calculé et divisé par le nombre total d'individus afin d'obtenir des proportions. La distribution de probabilités est présentée dans un graphique comme celui de la figure 10.1, à la page suivante, où l'on trouve en abscisse les classes et le nombre d'individus par classe, et les proportions correspondantes en ordonnée.

EXEMPLE 10.14 (suite)**FIGURE 10.1 ■ Les valeurs de la pression artérielle diastolique****La distribution normale**

En statistique, une variable se distribue normalement si elle possède certaines caractéristiques mathématiques : être continue, en forme de cloche et symétrique par rapport à la moyenne (μ), comme l'illustre la figure 10.2 (cholestérolémie).

FIGURE 10.2 ■ Une illustration d'une distribution normale (cholestérolémie)

Le premier écart type (σ) correspond à la distance entre la moyenne et le point d'inflexion de la courbe, c'est-à-dire l'endroit où la courbe, de convexe qu'elle était, devient concave. La moyenne et l'écart type sont les deux paramètres de la **distribution normale** : ils constituent les deux seuls éléments d'information nécessaires pour définir celle-ci. Il y a autant de courbes normales qu'il y a de paramètres μ et σ différents. Comme la distribution normale est une distribution de probabilités, l'aire sous la courbe est égale à 1, car elle représente la probabilité que tous les événements possibles surviennent.

D'autres caractéristiques de la distribution normale sont également importantes. D'une part, la moyenne, la médiane et le mode ont une même valeur. D'autre part, la moyenne ± 1 écart type contient approximativement 68,3 % de l'aire sous la courbe, soit 68,3 % des événements possibles ; la moyenne ± 2 écarts types contient

approximativement 95,4 % des événements possibles ; et la moyenne ± 3 écarts types contient approximativement 99,7 % des événements possibles.

EXEMPLE 10.15

Si les valeurs possibles de la pression artérielle diastolique chez les 1 000 individus de la ville de Québec présentées à la figure 10.1 sont distribuées normalement, la pression artérielle se situera entre -1 écart type et $+1$ écart type (70 mm Hg à 94 mm Hg) chez approximativement 683 individus, de 55 mm Hg à 109 mm Hg chez 954 individus et de 50 mm Hg à 114 mm Hg chez 997 individus.

Il faut noter qu'une certaine prudence s'impose lorsqu'on emploie l'expression « normale », car le qualificatif « normal » est d'usage fréquent. Ici, l'expression ne désigne pas la normalité clinique, c'est-à-dire ce que l'on observe chez des individus normaux ou sains, mais plutôt la normalité statistique.

EXEMPLE 10.16

Lors d'une récente réunion d'une association professionnelle québécoise, son président explique que, à l'échelle canadienne, les salaires suivent une distribution « normale » avec une moyenne de 70 000 \$ par an. Pourtant, dans la salle, plusieurs personnes ne trouvent pas « normal » de toucher un salaire moins élevé que celui de collègues travaillant dans la province voisine. Ainsi, bien que le salaire annuel de 70 000 \$ puisse, à l'échelle canadienne, se situer dans la moyenne statistique pour la profession, il n'est pas forcément comparable avec celui que touchent les participants à la réunion. Qu'est-ce qui est normal ? Qu'est-ce qui est anormal ? Il n'est pas contradictoire de constater que les salaires se distribuent normalement à l'échelle canadienne, tout en considérant comme anormal le fait qu'ils soient bas ou élevés dans des régions données.

La distribution normale centrée et réduite

La distribution normale est utile pour évaluer la proportion des éléments qui se situent dans un intervalle de valeurs de la variable étudiée. Ses caractéristiques mathématiques correspondent à un nombre entier d'écarts types ; cependant, ceux-ci ne sont pas les seuls intervalles intéressants. Ainsi, si l'on s'intéresse à la proportion d'individus dont la pression artérielle diastolique excède 80 mm Hg, ou à la proportion d'individus chez qui cette valeur est inférieure à 70 mm Hg, il devient impossible d'obtenir les proportions sans effectuer un calcul très difficile². Pour apprécier ces proportions, il est nécessaire d'utiliser une distribution où les proportions ont été calculées à l'avance pour des intervalles plus fins.

Comme nous l'avons signalé plus haut, il y a autant de courbes de distribution normale qu'il y a de types de variables et de répartitions de celles-ci ; ce principe est illustré dans les figures 10.3 et 10.4, qui représentent respectivement la distribution des différentes valeurs de la cholestérolémie totale et de la pression artérielle systolique dans la population.

Afin d'éviter d'avoir une distribution différente pour chaque valeur de la moyenne et de l'écart type, on utilise la **distribution normale centrée et réduite** (voir la figure 10.5). Il s'agit d'une distribution normale pour laquelle la moyenne est de 0 et l'écart type de 1. Cette distribution est aussi appelée **distribution standard** ou **distribution de Z**.

FIGURE 10.3 ■ La distribution des valeurs de la cholestérolémie totale dans la population

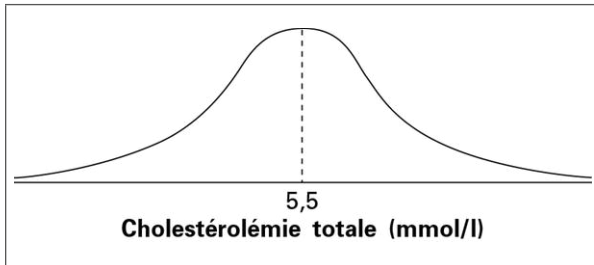


FIGURE 10.4 ■ La distribution des valeurs de la pression artérielle systolique dans la population

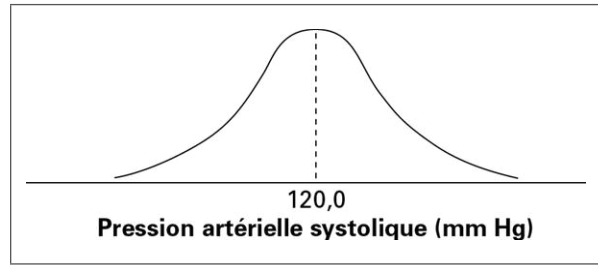
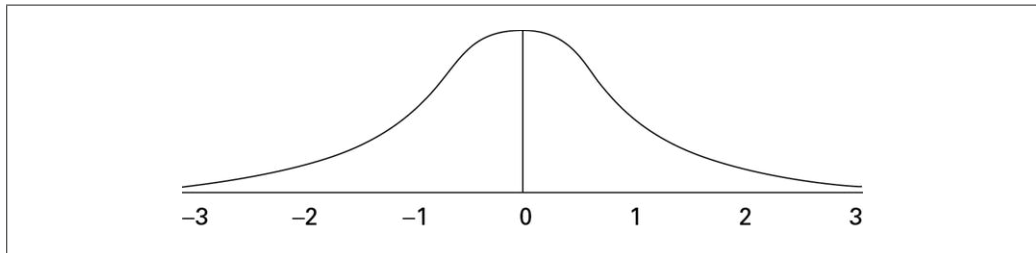


FIGURE 10.5 ■ La distribution normale centrée et réduite



On peut transformer l'ensemble des distributions normales en une distribution normale centrée et réduite en effectuant l'opération suivante (où Z représente une variable ayant une distribution normale centrée et réduite et x , la variable à transformer) : on appelle communément cette opération le **score Z** .

$$Z = \frac{x - \mu}{\sigma}$$

EXEMPLE 10.17

Quelle est la proportion des individus de la population chez qui les valeurs de la pression artérielle diastolique sont supérieures à 80 mm Hg?

Si l'on sait que la moyenne de la pression artérielle diastolique dans la population est de 78 mm Hg et que l'écart type est de 2 mm Hg, le score Z s'exprime comme suit :

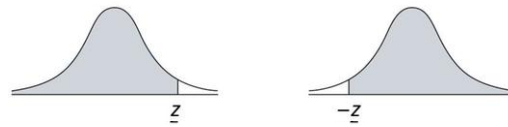
$$Z = \frac{80 - 78}{2}, \text{ soit } 1.$$

Pour obtenir la proportion d'individus se situant dans n'importe quelle aire sous la courbe, on utilise généralement une table comme celle présentée dans le tableau 10.4.

Le tableau 10.4 indique que la probabilité correspondant à $Z = 1,00$ est de 0,8413. C'est la probabilité que la pression artérielle diastolique soit inférieure à 80 mm Hg (zone ombrée de la figure). Comme on s'intéresse à la proportion des personnes qui ont une pression artérielle supérieure à 80 mm Hg dans la population, il faut

soustraire cette probabilité de 1 ($1 - 0,8413$). La proportion de la population présentant une pression artérielle diastolique supérieure à 80 mm Hg est donc de 15,9 %.

TABLEAU 10.4 ■ Une table de la distribution normale centrée et réduite



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8909	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9321	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9976	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986



TABLEAU 10.4 ■ Une table de la distribution normale centrée et réduite (suite)

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,00	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,10	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,20	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,30	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,40	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,50	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,60	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,70	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,80	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Source : Adapté de Daniel, W. *Biostatistics: A Foundation for Analysis in the Health Sciences*, 4^e édition, New York, John Wiley, 1989, p. 687-688.

Le théorème de la limite centrale

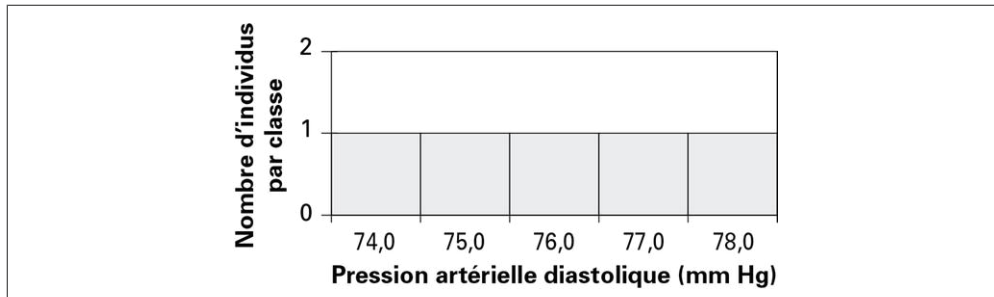
La distribution normale est un outil essentiel en statistique. Malheureusement, très peu de variables sont dotées de toutes les caractéristiques mathématiques de cette distribution. On peut contourner ce problème en utilisant ce que l'on appelle la **distribution échantillonnale** ou **distribution d'échantillonnage**. Celle-ci diffère de la distribution de probabilités des variables. Il s'agit de la distribution de probabilités d'une statistique pour tous les échantillons de même taille pouvant être formés à partir d'une population. Pour définir la distribution échantillonnale, on considère, outre la statistique (soit, dans le cas présent, la moyenne), la sélection de l'échantillon, la taille de celui-ci et la population à l'étude.

EXEMPLE 10.18

Supposons que la population étudiée par Lisa et présentée dans le tableau 10.1 ne compte que les individus 1 à 5, chez qui les valeurs de la pression artérielle diastolique (mm Hg) sont les suivantes: individu 1: 74,0; individu 2: 75,0; individu 3: 76,0; individu 4: 77,0; individu 5: 78,0.

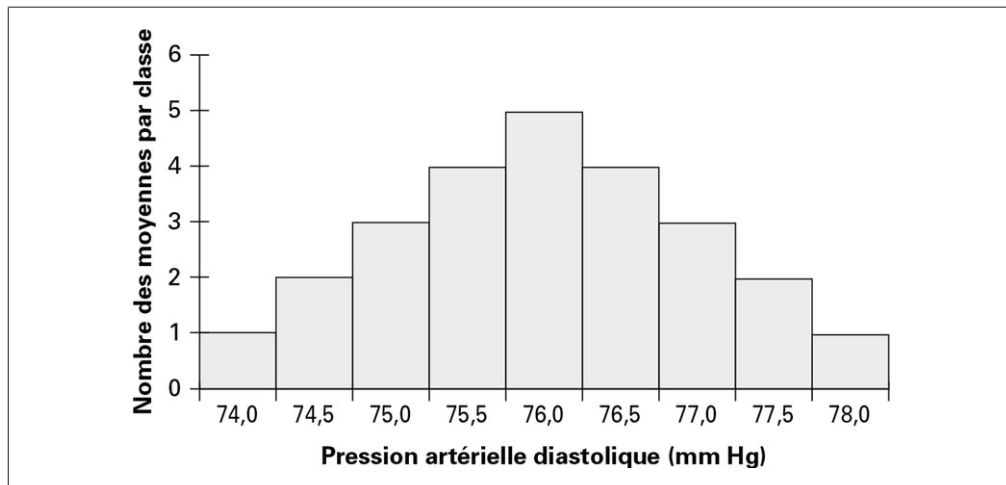
La moyenne des valeurs de la pression artérielle diastolique est de 76 mm Hg et l'écart type est de 1,4 mm Hg. La distribution des valeurs de la variable dans la population étudiée représentée à la figure 10.6 n'est donc pas normale, car la variable n'est pas dotée de toutes les caractéristiques mathématiques de la distribution normale.

Or, à partir de cette population, on peut créer une **distribution des moyennes échantillonnales**. En considérant par exemple tous les échantillons possibles de deux individus (le même individu pouvant être choisi plus d'une fois), on obtient la totalité des échantillons potentiels pour lesquels on peut calculer la moyenne de la pression artérielle diastolique. Ainsi, 25 échantillons de 2 individus peuvent être formés à partir de la population de 5 individus. Le tableau 10.5 présente ces échantillons et les moyennes échantillonnales correspondantes.

EXEMPLE 10.18 (suite)**FIGURE 10.6 ■ La distribution des valeurs de la pression artérielle diastolique****TABLEAU 10.5 ■ Les échantillons pouvant être formés**

Échantillon	Individus sélectionnés	Pression artérielle diastolique (mm Hg)	Moyenne (mm Hg)	Échantillon	Individus sélectionnés	Pression artérielle diastolique (mm Hg)	Moyenne (mm Hg)
1	1; 1	74,0; 74,0	74,0	14	3; 4	76,0; 77,0	76,5
2	1; 2	74,0; 75,0	74,5	15	3; 5	76,0; 78,0	77,0
3	1; 3	74,0; 76,0	75,0	16	4; 1	77,0; 74,0	75,5
4	1; 4	74,0; 77,0	75,5	17	4; 2	77,0; 75,0	76,0
5	1; 5	74,0; 78,0	76,0	18	4; 3	77,0; 76,0	76,5
6	2; 1	75,0; 74,0	74,5	19	4; 4	77,0; 77,0	77,0
7	2; 2	75,0; 75,0	75,0	20	4; 5	77,0; 78,0	77,5
8	2; 3	75,0; 76,0	75,5	21	5; 1	78,0; 74,0	76,0
9	2; 4	75,0; 77,0	76,0	22	5; 2	78,0; 75,0	76,5
10	2; 5	75,0; 78,0	76,5	23	5; 3	78,0; 76,0	77,0
11	3; 1	76,0; 74,0	75,0	24	5; 4	78,0; 77,0	77,5
12	3; 2	76,0; 75,0	75,5	25	5; 5	78,0; 78,0	78,0
13	3; 3	76,0; 76,0	76,0				

La moyenne des moyennes échantillonnales est de 76 mm Hg, ce qui correspond à la moyenne calculée dans la population. L'écart type des moyennes échantillonnales est de 1 mm Hg, ce qui s'avère inférieur à l'écart type calculé dans la population. La distribution des valeurs des moyennes échantillonnales, dont la figure 10.7, à la page suivante, constitue un exemple, a une forme qui rappelle celle de la distribution normale.

EXEMPLE 10.18 (suite)**FIGURE 10.7 ■ La distribution des valeurs des moyennes échantillonnales**

L'écart type des moyennes échantillonnales est aussi connu sous le nom **d'erreur type**. En pratique, on ne connaît pas l'erreur type. On doit l'estimer à partir d'un échantillon par le calcul s/\sqrt{n} , où s représente l'écart type de l'échantillon, et n la taille de l'échantillon. L'erreur type ne fournit aucune information sur la dispersion de la variable dans l'échantillon sélectionné à partir de la population à l'étude, au contraire de l'écart type. Elle indique avec quel degré de certitude la moyenne échantillonnale estime la moyenne dans la population. Ainsi, l'erreur type pour l'échantillon formé des individus 1 et 3 présentés dans le tableau 10.1 est de 1 mm Hg. L'erreur type d'un autre échantillon formé des individus 1, 2 et 3 est de 0,58 mm Hg. L'erreur type calculée dans l'échantillon formé de trois individus est plus petite que celle calculée dans l'échantillon composé de deux individus. La moyenne de l'échantillon formé de trois individus estime donc la moyenne de la population avec un plus haut degré de certitude. En fait, plus on augmente la taille de l'échantillon, mieux on estime la valeur du paramètre.

Dans une population dont la moyenne d'une variable est μ et l'écart type est σ , la distribution de probabilités des moyennes échantillonnales a les propriétés suivantes.

- La moyenne de la distribution échantillonnale, c'est-à-dire la moyenne des moyennes échantillonnales, est égale à la moyenne de la variable dans la population (soit le paramètre).
- L'erreur type est l'écart type des moyennes échantillonnales et peut être estimée par s/\sqrt{n} .
- Si la distribution de la variable est normale, la distribution des moyennes échantillonnales l'est aussi. De plus, pour une taille d'échantillon suffisamment grande ($n \geq 30$), la distribution des moyennes échantillonnales est normale, quelle que soit la distribution de la variable dans la population.

Ces propriétés sont à la base d'un important théorème en statistique, connu sous le nom de **théorème de la limite centrale**. Ce théorème illustre l'importance de la loi de distribution normale, qui peut être utilisée même lorsque les variables ne sont pas distribuées normalement; il permet d'effectuer l'inférence statistique. Il est donc possible de transformer la moyenne en une variable de distribution normale centrée et réduite. Dans ce cas, la transformation Z , ou distribution de Z , s'exprime de la façon suivante.

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

EXEMPLE 10.19

Lorsque la moyenne de la pression artérielle diastolique dans une population est de 76 mm Hg et que l'écart type dans l'échantillon est de 10 mm Hg, quelle est la probabilité qu'un échantillon de 25 individus sélectionnés dans cette population présente une moyenne de la pression artérielle diastolique supérieure à 80 mm Hg?

Le résultat de la transformation des données sera donc :

$$Z = \frac{80 - 76}{10/\sqrt{25}}, \text{ soit } 2.$$

Le tableau 10.4 permet de calculer que 2,28 % des éléments se situent dans l'intervalle au-dessus de 2, étant donné que la probabilité correspondant à $Z = 2$ est de 0,9972. Par conséquent, 2,28 % ($1 - 0,9972 = 0,0228$) des échantillons tirés de cette population présenteront une moyenne de la pression artérielle diastolique supérieure à 80 mm Hg.

Cette opération de transformation trouvera une application utile au moment de l'utilisation des intervalles de confiance (*voir à ce sujet le chapitre 11*).

À retenir

- La moyenne et l'écart type sont les deux seuls éléments d'information nécessaires pour définir la distribution normale.
- La distribution normale centrée et réduite est une distribution normale avec une moyenne de 0 et un écart type de 1.
- La distribution échantillonnale consiste en la distribution de probabilités d'une statistique pour tous les échantillons de même taille pouvant être formés à partir d'une population.
- L'écart type informe sur la dispersion de la variable dans la population (σ) ou dans l'échantillon (s), alors que l'erreur type informe sur la précision de l'estimation de la moyenne dans la population. Plus l'erreur type est petite, meilleure est la précision de l'estimation.
- Le théorème de la limite centrale permet d'effectuer l'inférence statistique, même lorsque les variables ne sont pas distribuées normalement.

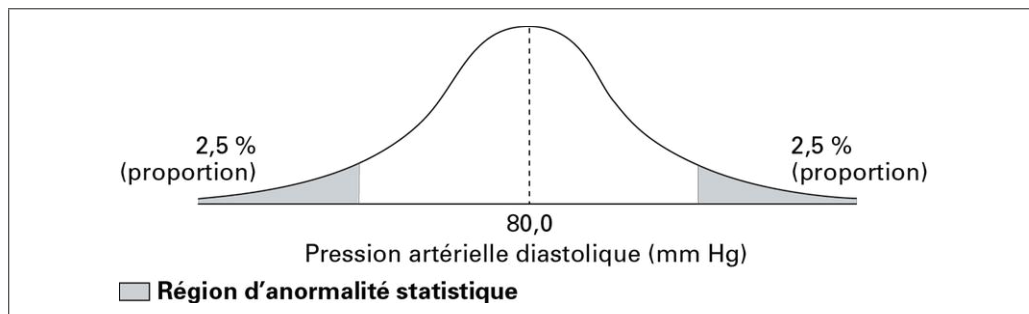
10.4 La normalité statistique et la normalité clinique

Il est essentiel de faire la distinction entre la normalité statistique, qui renvoie à la répartition des valeurs observées pour une variable dans une population, et la normalité clinique, qui pourrait correspondre, par exemple, au seuil à partir duquel un résultat d'examen médical témoigne d'un risque pour la santé d'une personne.

10.4.1 La normalité statistique

Comme il a été mentionné au début du présent chapitre, Lisa se demande si l'anomalie de la pression artérielle que présente Charles s'observe fréquemment. Pour répondre à cette question, un échantillon peut être sélectionné, et l'on peut observer sa distribution de probabilités. La pression artérielle de Charles s'avère anormale sur le plan statistique si elle se situe à l'une ou l'autre des extrémités de la distribution. Dans ce cas, Charles sera considéré comme « anormal » sur le plan statistique. Pour en arriver à cette conclusion, il faut déterminer le niveau extrême : celui-ci définit la région critique, c'est-à-dire la région dite « d'anormalité statistique ». On considère souvent que la région critique se situe dans la dernière tranche de 2,5 % de chaque côté de la courbe de distribution, comme le représente la figure 10.8 ; la région critique contient alors 5 % des éléments observés.

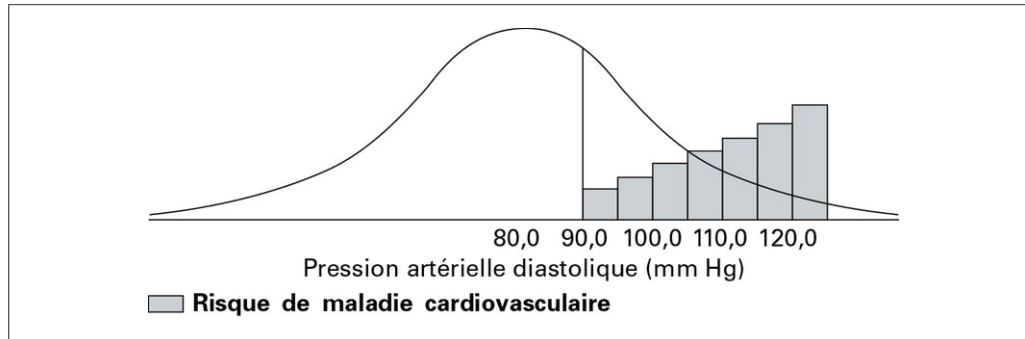
FIGURE 10.8 ■ La normalité statistique



10.4.2 La normalité clinique

Lisa se demande également si le niveau de pression artérielle que présente Charles constitue un risque pour sa santé. La normalité statistique ne permet pas de répondre à cette question, car on ne connaît pas, à partir d'un échantillon d'individus dont on a mesuré la pression artérielle, le risque de complications qui peut être associé aux différentes valeurs de celle-ci. Certaines études ont toutefois tenté de faire le lien entre les différentes valeurs de la pression artérielle et le risque de maladie cardiovasculaire ou d'accident vasculaire cérébral. La plupart de ces études indiquent que les individus dont la pression artérielle diastolique est égale ou supérieure à 90 mm Hg présentent un risque plus élevé de maladie cardiovasculaire que ceux dont la pression artérielle diastolique se situe à un niveau inférieur à 90 mm Hg, comme l'illustre la figure 10.9.

La normalité clinique correspond à des valeurs de la pression artérielle diastolique inférieures à 90 mm Hg. Lisa peut donc en conclure que la pression artérielle mesurée chez Charles (156/100 mm Hg) constitue un risque pour sa santé, puisqu'elle est élevée.

FIGURE 10.9**■ La normalité clinique : augmentation du risque de maladie cardiovasculaire****En résumé**

La démarche statistique s'applique à plusieurs types de variables, peu importe leur distribution de probabilités. L'inférence statistique peut être utilisée pour analyser les données obtenues lors d'études sur des échantillons. Il faut cependant s'assurer de la validité externe des résultats des

études. L'interprétation des résultats obtenus à partir de recherches s'effectuera à la lumière d'autres processus faisant partie de la démarche statistique et sur lesquels portent les deux prochains chapitres.

Exercices

Voir la section complète des exercices sur le Web.

- Pour vous assurer de constituer un échantillon représentatif d'une population en réduisant le risque de biais de sélection, quel type d'échantillon devez-vous privilégier?
 - Électif.
 - Aléatoire.
 - Apparié.
- Laquelle de ces deux mesures nous informe sur la précision de l'estimation de la moyenne de la population?
 - L'écart type.
 - L'erreur type.
- Parmi les énoncés suivants, lequel définit une probabilité objective?
 - Une probabilité définie de façon intuitive, sur la base d'une expérience personnelle ou d'une impression basée sur l'expérience.
 - Une probabilité établie sur la base d'une expérience répétée un très grand nombre de fois.
- Vrai ou faux. La validité externe d'une étude renvoie à la capacité de généraliser un résultat obtenu à partir d'un échantillon à une population cible.
- Dans un échantillon où la répartition selon l'âge se distribue selon les caractéristiques de la courbe normale, où la moyenne d'âge est de 40 ans et où l'écart type de l'âge est de 3 ans, quelle est la probabilité qu'un individu choisi au hasard soit âgé de 37 à 43 ans?
- Dans un échantillon où la moyenne d'âge et la médiane d'âge sont de 40 ans, quelle est la probabilité qu'un individu choisi au hasard soit âgé de moins de 40 ans?

Pour les réponses, voir le corrigé à la fin du manuel.

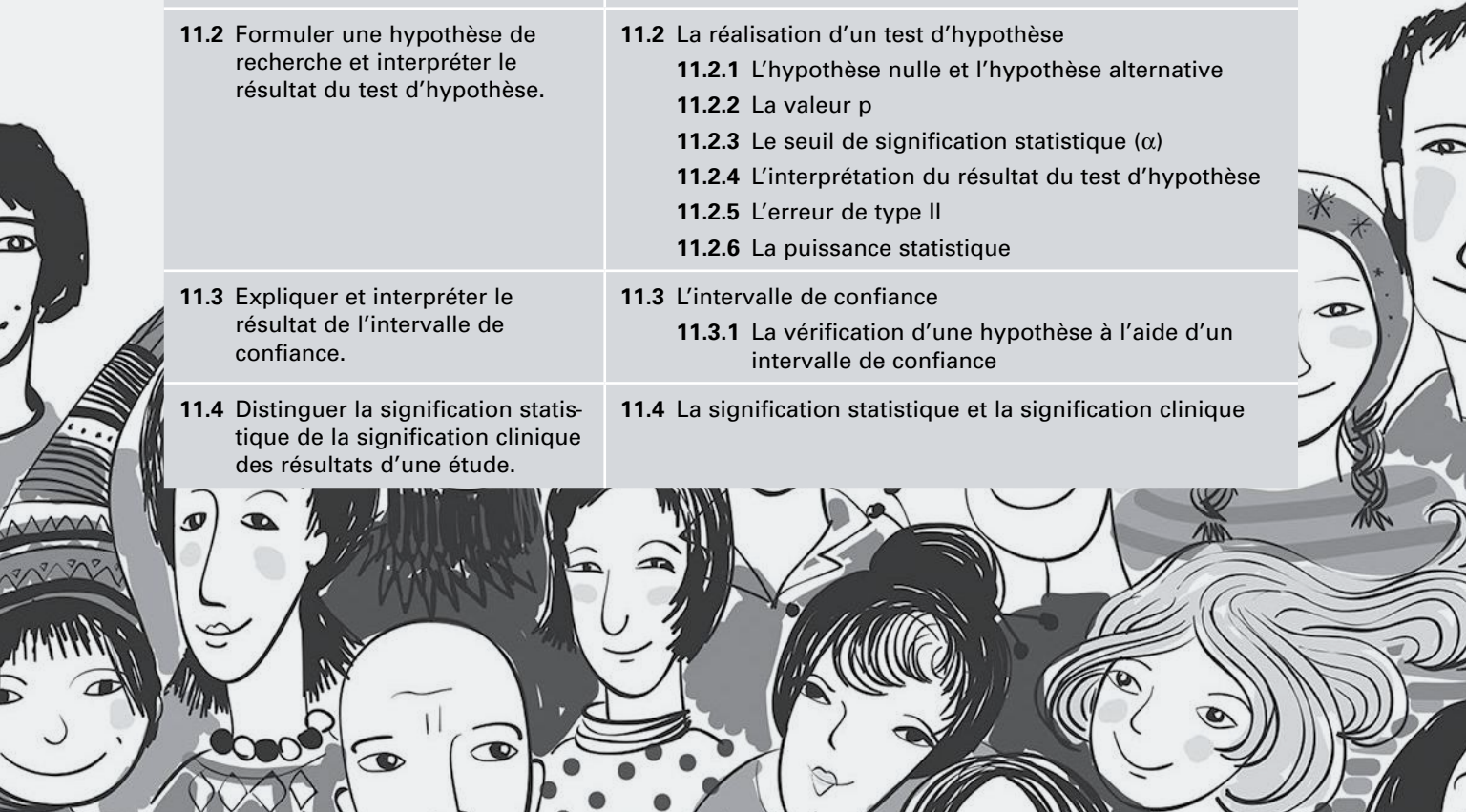
La signification statistique et la signification clinique

Collaboration à la 3^e édition :
Suzanne Gingras

Au terme de ce chapitre, vous pourrez :

- expliquer les concepts et les méthodes de base de la signification statistique ;
- juger de la validité scientifique d'une étude et d'un résultat ;
- juger de l'utilité du contenu et des résultats d'un écrit scientifique au regard de votre pratique ;
- prendre une décision éclairée sur la pertinence de modifier votre pratique clinique à la lumière des résultats d'études scientifiques.

Objectifs spécifiques	Contenu
11.1 Exposer le raisonnement qui sous-tend l'inférence statistique.	11.1 La signification statistique 11.1.1 L'interprétation des résultats de l'inférence statistique 11.1.2 Les limites de l'échantillonnage
11.2 Formuler une hypothèse de recherche et interpréter le résultat du test d'hypothèse.	11.2 La réalisation d'un test d'hypothèse 11.2.1 L'hypothèse nulle et l'hypothèse alternative 11.2.2 La valeur p 11.2.3 Le seuil de signification statistique (α) 11.2.4 L'interprétation du résultat du test d'hypothèse 11.2.5 L'erreur de type II 11.2.6 La puissance statistique
11.3 Expliquer et interpréter le résultat de l'intervalle de confiance.	11.3 L'intervalle de confiance 11.3.1 La vérification d'une hypothèse à l'aide d'un intervalle de confiance
11.4 Distinguer la signification statistique de la signification clinique des résultats d'une étude.	11.4 La signification statistique et la signification clinique





► Mise en situation

Charles se présente à son rendez-vous à la polyclinique Milo pour le suivi de sa pression artérielle. Il a bien observé les consignes de Lisa, l'omnipraticienne, mais sa pression artérielle demeure élevée, se situant à 160/100 mm Hg. Les résultats des analyses de laboratoire révèlent par ailleurs une cholestérolémie totale légèrement élevée. Lisa a posé un diagnostic d'hypercholestérolémie et d'hypertension artérielle (HTA) essentielle. Parmi les conséquences possibles de ces problèmes de santé figure l'augmentation du risque de maladie cardiovasculaire. Certains médicaments permettent de traiter efficacement l'HTA, mais le traitement doit se poursuivre sur une très longue période. Lisa prescrit un antihypertenseur à Charles et lui recommande une diète pauvre en sel ainsi que le maintien de saines habitudes de vie. Elle lui suggère aussi de consulter Stéphanie, la nutritionniste de la polyclinique, afin d'établir les habitudes alimentaires susceptibles de réduire

son hypercholestérolémie. Lisa prévoit revoir Charles un mois plus tard.

Pertinence clinique

Lisa a décelé chez Charles la présence de facteurs associés à une augmentation du risque de maladie cardiovasculaire. Elle a informé son patient des conséquences qu'entraîne l'HTA pour sa santé et lui a prescrit un traitement jugé efficace. Ces conduites constituent autant de facettes du travail clinique qui s'appuient sur des données particulières. Celles-ci sont obtenues à partir d'études épidémiologiques. Le plus souvent, ces études portent sur un échantillon, c'est-à-dire sur un sous-ensemble de la population. La question suivante se pose donc : « Le résultat d'une étude, obtenu à partir d'un échantillon, constitue-t-il une estimation valable de celui que l'on aurait obtenu si l'on avait observé l'ensemble des individus de la population d'où provient cet échantillon? »

Le présent chapitre est en continuité avec le précédent. Il traite de l'inférence statistique, c'est-à-dire du processus permettant d'induire la valeur d'un paramètre à partir d'une mesure obtenue (soit une statistique) dans un échantillon représentatif de la population. On y aborde aussi les critères de qualité et de validité à respecter pour ne pas être induit en erreur lorsque nous généralisons les résultats tirés d'un échantillon et devons juger de la pertinence de modifier notre pratique clinique à la lumière de ces résultats. Nous ferons alors une distinction entre la signification statistique et la signification clinique.

11.1 La signification statistique

Lisa vient de lire une publication qui présente les résultats d'une étude portant sur un nouveau médicament contre l'HTA. De sa lecture attentive, elle retient les éléments suivants.

L'étude consiste en un essai thérapeutique visant à évaluer les avantages présumés d'un nouveau médicament antihypertenseur, le Novel. Quatre-vingts hommes blancs hypertendus (pression artérielle diastolique ≥ 90 mm Hg en l'absence d'un traitement), âgés de 35 ans à 49 ans, sont répartis aléatoirement (par hasardisation) en 2 groupes. Les 40 sujets du premier groupe, appelé «groupe N», reçoivent le médicament Novel; les sujets du second groupe, appelé «groupe P», reçoivent le plus efficace des médicaments actuellement en usage, le Périn. Le nouveau médicament a été évalué sous deux aspects, soit l'efficacité et la rapidité d'action: l'efficacité, en comparant la moyenne des valeurs de la pression artérielle diastolique (PAD) des deux groupes, et la rapidité d'action, en comparant les proportions d'individus de chacun de ces deux groupes présentant une PAD normale après deux mois de traitement. Ni les sujets ni les chercheurs ne savaient à quel groupe appartenait chacun des individus (il s'agissait d'une étude en double aveugle). Tous les sujets ont été suivis pendant la durée complète de l'étude.

À la fin de la période d'observation, la moyenne de la PAD pour le groupe N (\bar{x}_n) est de 86 mm Hg, alors qu'elle est de 89 mm Hg pour le groupe P (\bar{x}_p), ce qui constitue un écart observé de -3 mm Hg entre les deux groupes. Ce résultat est jugé statistiquement significatif pour un seuil de signification fixé à 0,05 (5%). Par ailleurs, 80% ($p_n = 32/40$) des individus du groupe N présentent une PAD normale à la fin de la période d'observation, tandis que ce pourcentage est de 65% ($p_p = 26/40$) pour le groupe P, ce qui constitue un écart de 15%. Ce résultat n'est pas jugé statistiquement significatif pour un seuil de signification fixé à 0,05. Les auteurs concluent à la plus grande efficacité du nouveau médicament, mais non à une plus grande rapidité d'action. Cependant, que veut dire l'expression «statistiquement significatif»?

11.1.1 L'interprétation des résultats de l'inférence statistique

Tout d'abord, l'étude est de type expérimental hasardisé. Elle a été effectuée en double aveugle. Ce type d'étude réduit au minimum la possibilité de biais et conduit à des résultats en général valides. Ainsi, l'étude a une bonne validité interne. La validité externe concerne, quant à elle, la capacité de généraliser un résultat à une population cible. Ici, la population à l'étude est constituée d'hommes blancs hypertendus âgés de 35 ans à 49 ans. Par conséquent, on ne pourrait appliquer les résultats de cette étude aux hommes âgés de moins de 35 ans ou de plus de 49 ans, aux femmes ou aux personnes noires, par exemple.

Les résultats de cette étude ont été obtenus à partir de l'observation de 2 échantillons de 40 hommes hypertendus*. En fonction de ces échantillons, on peut conclure à la plus grande efficacité du nouveau médicament, puisqu'on observe un écart de -3 mm Hg entre la moyenne de la PAD du groupe N et celle du groupe P. Par contre, ce qui

* Pour la suite du chapitre, l'expression «hommes hypertendus» désignera toujours des hommes blancs hypertendus âgés de 35 ans à 49 ans.

intéresse vraiment Lisa, c'est de juger de l'efficacité de ce nouvel antihypertenseur chez tous les hommes hypertendus susceptibles de le recevoir. L'étude qu'elle consulte donne-t-elle la possibilité d'y arriver à partir de la seule observation de 2 échantillons de 40 hommes hypertendus? En somme, Lisa doit savoir utiliser les résultats de cette étude pour appuyer ou non sa décision clinique.

Cette question relève de l'inférence statistique, c'est-à-dire du processus qui permet d'induire la valeur d'un paramètre à partir d'une mesure obtenue (soit une statistique) dans un échantillon représentatif de la population. L'inférence statistique peut être effectuée selon deux approches : le test d'hypothèse¹ et l'intervalle de confiance. Il faut toutefois demeurer prudent : les méthodes que sous-tendent ces approches ne peuvent suppléer à un protocole d'étude déficient et ne permettent pas de porter un jugement sur les biais susceptibles de rendre moins valides les résultats obtenus. Il ne faut pas confondre validité de l'étude et inférence statistique.

Les deux approches statistiques susmentionnées comportent des aspects mathématiques relativement complexes qu'il n'est pas nécessaire de maîtriser pour comprendre la logique qui sous-tend leur utilisation. Pour un professionnel de la santé, il s'agit plutôt d'être capable de reconnaître et d'interpréter correctement les informations d'ordre statistique. L'énoncé « ce résultat est statistiquement significatif pour un seuil de signification fixé à 0,05 » est fréquent dans les publications scientifiques. Il s'agit tout simplement de la conclusion d'un test d'hypothèse... mais de quelle hypothèse?

À retenir

- L'inférence statistique est un processus qui permet d'induire la valeur d'un paramètre à partir d'une mesure obtenue dans un échantillon représentatif de la population.
- Le test d'hypothèse et l'intervalle de confiance sont deux approches statistiques qui relèvent de l'inférence statistique.

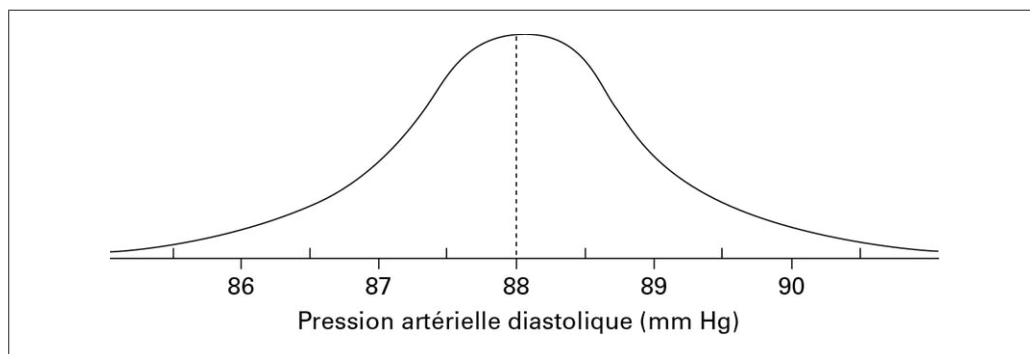
11.1.2 Les limites de l'échantillonnage

Dans l'article qui intéresse Lisa, les chercheurs ont émis l'idée selon laquelle l'efficacité du nouveau médicament antihypertenseur, le Novel, est différente de celle du médicament actuellement en usage, le Périn. Cela constitue leur hypothèse de recherche. Les chercheurs ont donc voulu vérifier cette hypothèse dans la population des hommes hypertendus. Puisqu'il est à peu près impossible d'observer la totalité de cette population, l'étude a porté sur 2 échantillons de 40 individus. À la lumière de cette seule expérience, la différence observée (-3 mm Hg) a été jugée statistiquement significative et les chercheurs ont conclu à la plus grande efficacité du Novel dans la population des hommes hypertendus.

Pour comprendre le raisonnement statistique qui appuie cette conclusion, revoyons le déroulement de cette étude, mais de façon à utiliser les propriétés du théorème de la limite centrale dont il a été question dans le chapitre 10 (*voir la page 212*).

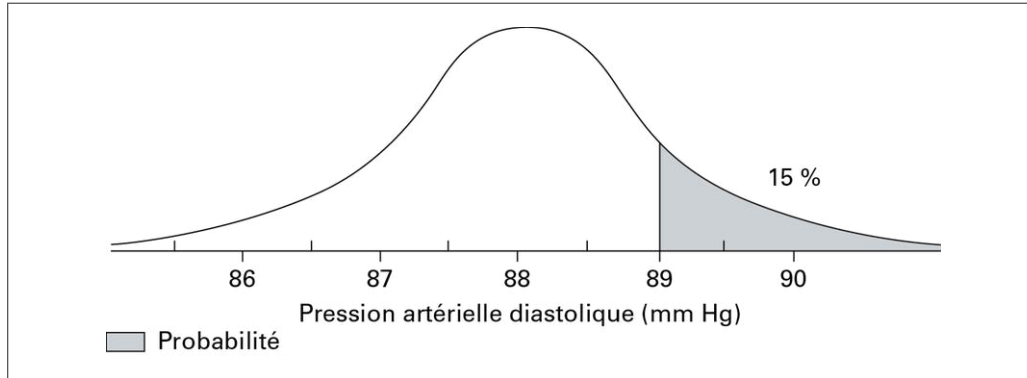
Les chercheurs ont réalisé une seule expérience avec 2 échantillons de 40 hommes hypertendus. Après 2 mois de traitement, ils ont observé une moyenne de la PAD de 89 mm Hg dans l'échantillon des 40 hommes hypertendus traités avec le Périn. Ce résultat est une estimation de la valeur réelle de la moyenne de la PAD qu'on pourrait obtenir si l'on observait tous les hommes hypertendus traités avec ce médicament. Si l'on répétait cette expérience, dans les mêmes conditions, à partir d'un autre échantillon de 40 hommes hypertendus traités avec le Périn, la moyenne de la PAD serait sans doute un peu différente du premier résultat obtenu, par exemple de 86 mm Hg. Une troisième expérience portant sur un autre échantillon de la même taille permettrait encore une fois de calculer une moyenne de la PAD différente, par exemple de 90 mm Hg. Ces différences entre les moyennes de la PAD mesurées sont attribuables aux fluctuations d'échantillonnage, au hasard en quelque sorte. Si l'on calculait les moyennes de la PAD pour tous les échantillons possibles de 40 hommes hypertendus traités avec le Périn (ce nombre est presque illimité), on observerait que les différentes valeurs obtenues suivent une distribution normale. La moyenne de cette distribution (soit la moyenne des moyennes) serait égale au paramètre (μ_p), c'est-à-dire à la valeur réelle de la moyenne de la PAD de cette population. Ce paramètre n'est pas connu, mais si l'on suppose, à titre d'exemple, que sa valeur dans la population des hommes hypertendus traités avec le Périn est de 88 mm Hg, on pourrait la représenter comme l'illustre la figure 11.1.

FIGURE 11.1 ■ La distribution d'échantillonnage des valeurs de la moyenne de la PAD (mm Hg) de tous les échantillons possibles de 40 hommes hypertendus traités avec le Périn (lorsque le paramètre supposé est de 88 mm Hg)



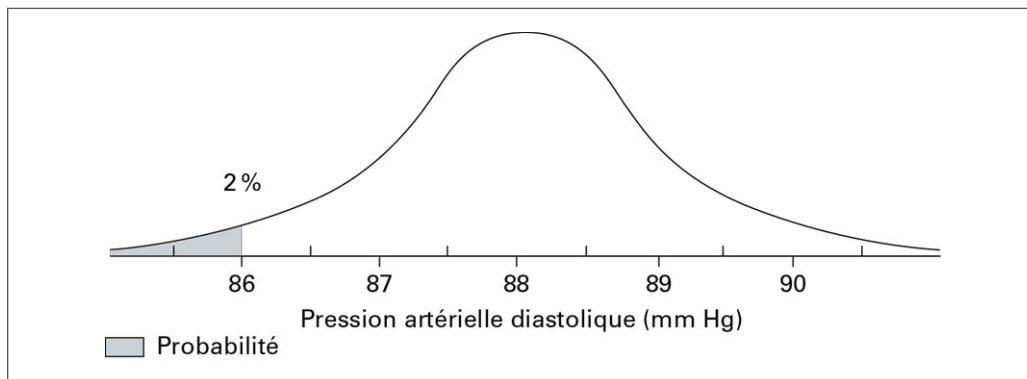
À partir de cette distribution, on remarque que certaines valeurs de la moyenne de la PAD sont plus ou moins fréquentes. La moyenne de la PAD de l'échantillon des hommes traités avec le Périn, soit 89 mm Hg, est plutôt rapprochée du paramètre supposé. En utilisant un échantillon de 40 individus sélectionnés dans cette population d'hommes hypertendus traités avec le Périn et en supposant que l'écart type y serait de 6 mm Hg, l'erreur type serait de $6/\sqrt{40}$, soit de 0,95; la probabilité d'obtenir une moyenne de la PAD de 89 mm Hg ou plus serait donc, selon la distribution normale centrée et réduite dont il a été question dans le chapitre 10, de 0,15 ou 15%. C'est ce qu'illustre la figure 11.2.

FIGURE 11.2 ■ La probabilité d'obtenir une moyenne de la PAD de 89 mm Hg ou plus à partir d'un échantillon de 40 individus sélectionnés dans une population d'hommes hypertendus traités avec le Périn (lorsque le paramètre supposé est de 88 mm Hg, et l'écart type dans l'échantillon, de 6 mm Hg)



Par ailleurs, dans cette même population d'hommes hypertendus traités avec le Périn, quelle serait la probabilité d'obtenir une moyenne échantillonnale de 86 mm Hg ou moins? Cette probabilité est de 0,02, soit de 2 %³, et elle est beaucoup plus faible que celle d'obtenir une moyenne de la PAD de 89 mm Hg ou plus (voir la figure 11.3).

FIGURE 11.3 ■ La probabilité d'obtenir une moyenne de la PAD de 86 mm Hg ou moins à partir d'un échantillon de 40 individus sélectionnés dans une population d'hommes hypertendus traités avec le Périn (lorsque le paramètre supposé est de 88 mm Hg, et l'écart type dans l'échantillon, de 6 mm Hg)



Les chercheurs ont également étudié un échantillon de 40 individus traités avec le Novel. Après deux mois de traitement, on a noté chez eux une moyenne de la PAD de 86 mm Hg. Ce résultat observé à partir d'un seul échantillon est inférieur à la moyenne de la PAD mesurée à partir de l'échantillon des hommes hypertendus traités avec le Périn (89 mm Hg); ici, le résultat semble indiquer une plus grande efficacité du Novel. On a vu qu'il est également possible d'observer une moyenne de la PAD de 86 mm Hg ou moins dans la population traitée avec le Périn selon une probabilité

de 2%. Alors, la moyenne de la PAD calculée à partir d'un échantillon d'individus recevant le Novel est-elle réellement différente de celle des hommes recevant le Périn? Ou a-t-on tiré au hasard un échantillon provenant de l'extrémité gauche de la même distribution d'échantillonnage, c'est-à-dire celle de la population des hommes hypertendus traités avec le Périn? Naturellement, on ne connaît pas les paramètres des populations. On a supposé que les paramètres de la population des hommes hypertendus traités avec le Périn étaient connus afin de montrer qu'il existe une probabilité, même si elle est très faible, que la moyenne de la PAD du groupe traité avec le Novel soit une valeur possible de la moyenne de la PAD du groupe traité avec le Périn.

Ainsi, à la lumière d'une seule expérience portant sur 2 échantillons de 40 individus, l'écart de -3 mm Hg observé entre les deux moyennes de la PAD peut être attribuable tant au simple fruit du hasard qu'à l'efficacité réelle du Novel. Ce sont là des limites dues aux fluctuations d'échantillonnage.

11.2 La réalisation d'un test d'hypothèse

Comment savoir si les résultats obtenus sont attribuables au hasard ou à la réelle efficacité du Novel? On observe, à partir de l'échantillon, que le nouveau traitement Novel est plus efficace que le médicament actuellement en usage, le Périn; mais peut-on conclure que le Novel est plus efficace que le Périn pour la population des hommes blancs hypertendus âgés de 35 ans à 49 ans?

En fait, il est possible de le découvrir en faisant de l'inférence statistique à partir, entre autres, de l'approche du test d'hypothèse, une procédure qui permet de choisir entre deux hypothèses statistiques qui sont faites sur la population.

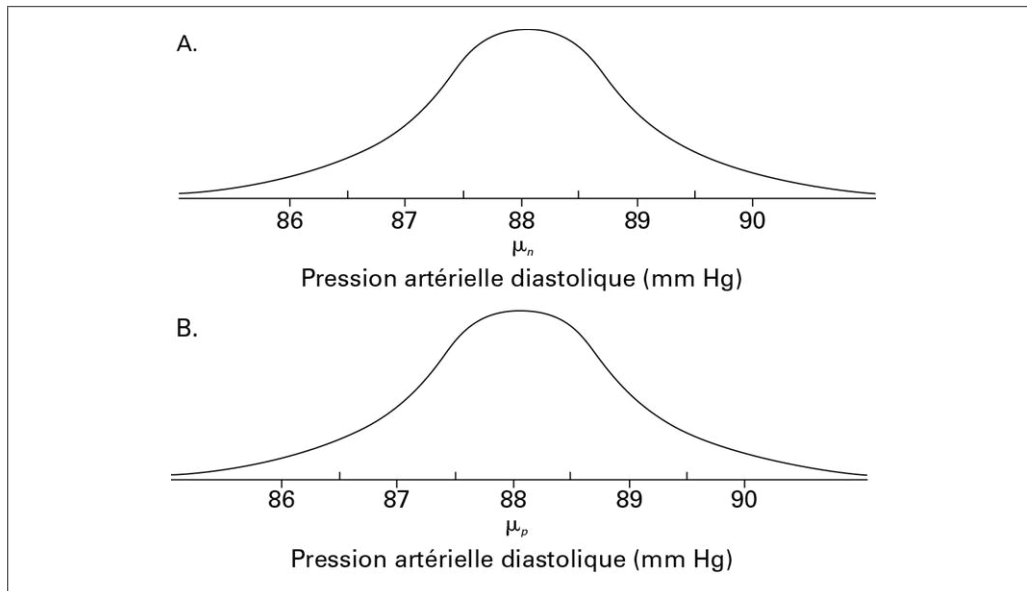
11.2.1 L'hypothèse nulle et l'hypothèse alternative

Pour tester une hypothèse de recherche, on confronte deux hypothèses statistiques: l'**hypothèse nulle**, qu'on désigne par H_0 , et l'**hypothèse alternative**, qu'on désigne par H_1 . Le test d'hypothèse est une approche générale qui s'applique à plusieurs types de mesures en épidémiologie: une différence de moyennes, une différence de proportions, des mesures d'association, etc. Dans le cas présent — une différence de moyennes —, l'hypothèse nulle consiste à supposer que, dans la population des hommes hypertendus, il n'y a pas de différence entre la moyenne de la PAD des individus soumis au Novel et celle des individus traités au Périn, une hypothèse qu'on exprime par $\mu_n = \mu_p$. L'hypothèse alternative correspond à l'hypothèse de recherche et prétend qu'il y a, dans cette population, une différence⁴ entre la moyenne de la PAD des individus soumis au Novel et celle des individus traités au Périn; on pourrait exprimer cela par $\mu_n \neq \mu_p$.

Après avoir formulé les hypothèses statistiques, on utilise une procédure qui vise à convaincre, au moyen du rejet de H_0 , de la vraisemblance de H_1 . Pour y parvenir, on suppose que H_0 est vraie. Si cette supposition ne résiste pas à l'examen des données des échantillons, elle est rejetée au profit de H_1 . Dans le cas présent, cette procédure amène à conclure qu'il y a, selon toute vraisemblance, une différence d'efficacité entre le Novel et le Périn.

Si H_0 est vraie, cela veut dire que les différences observées entre les deux groupes ne sont attribuables qu'aux fluctuations d'échantillonnage. En fait, cela signifie que les deux groupes proviennent de deux populations comparables en tous points — d'une même population, pourrait-on dire — en ce qui a trait à la moyenne de la PAD. La figure 11.4 représente ce phénomène. Par contre, si H_1 était vraie, les distributions d'échantillonnage des groupes A et B de la figure 11.4 seraient différentes, soit par leur valeur μ , soit par l'aplatissement des courbes.

FIGURE 11.4 ■ La distribution d'échantillonnage des valeurs de la moyenne de la PAD (mm Hg) de tous les échantillons possibles de 40 hommes hypertendus traités avec le Novel (en A) ou avec le Périn (en B), si H_0 est vraie ($\mu_n = \mu_p$)

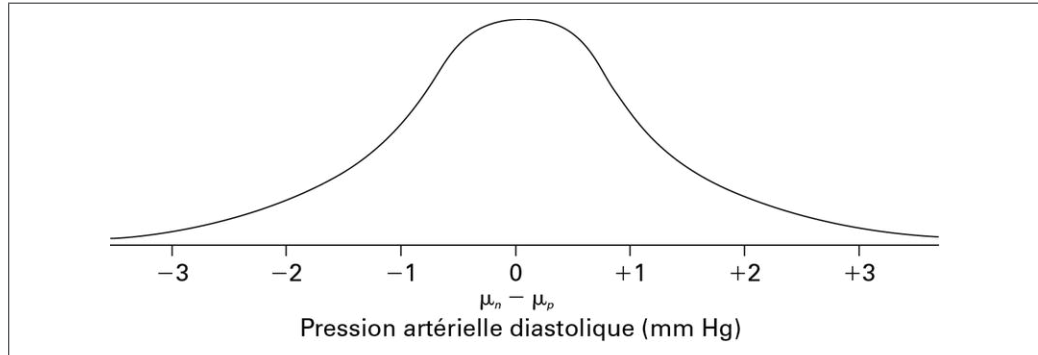


Ici, on ne compare pas une moyenne échantillonnale à un paramètre, mais deux moyennes échantillonnales entre elles. Si l'on répète un très grand nombre de fois l'expérience consistant à former 2 échantillons de 40 individus et à calculer chaque fois les écarts entre les moyennes de la PAD observées, on obtient des valeurs plus ou moins différentes de cet écart, certaines d'entre elles étant observées plus ou moins fréquemment. Il est alors possible de construire une distribution d'échantillonnage de ces écarts. Si H_0 est vraie, celle-ci sera approximativement normale et la plus grande proportion de ces mesures se situera autour de 0, la valeur moyenne de cette distribution⁵; la figure 11.5, à la page 226, propose une illustration de cette distribution. Plus rarement, toujours si H_0 est vraie, on observera des écarts positifs ou négatifs plus importants.

C'est à partir de cette distribution d'échantillonnage des écarts qu'il est possible de porter un jugement sur la vraisemblance de l'une ou l'autre des hypothèses. Le truc est simple: si H_0 est vraie, les valeurs des écarts gravitent autour de 0. Plus ces valeurs s'éloignent de 0, moins les écarts sont jugés compatibles avec H_0 . En présence d'un écart très important, il est permis de douter de la vraisemblance de H_0 . L'hypothèse nulle voulant qu'il n'y ait pas de différence entre les mesures de pression artérielle chez les hommes hypertendus traités avec le Novel comparativement à ceux traités avec le Périn est donc rejetée au

profit de l'hypothèse alternative jugée plus vraisemblable. On conclut alors que l'écart entre les moyennes de la PAD n'est probablement pas attribuable au hasard, à savoir aux fluctuations d'échantillonnage, mais bien à l'efficacité du nouveau médicament.

FIGURE 11.5 ■ La distribution d'échantillonnage des écarts entre les valeurs de la moyenne de la PAD (mm Hg) de tous les échantillons possibles de 40 hommes hypertendus traités avec le Novel ou le Périn, si H_0 est vraie ($\mu_n = \mu_p$)



À retenir

- Le test d'hypothèse donne de l'information sur le rôle que peut jouer le hasard dans l'observation des résultats obtenus à partir d'un échantillon de la population.
- Le test d'hypothèse est une démarche qui permet de faire, sur le plan statistique, le choix le plus judicieux entre deux hypothèses statistiques : l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1).

11.2.2 La valeur p

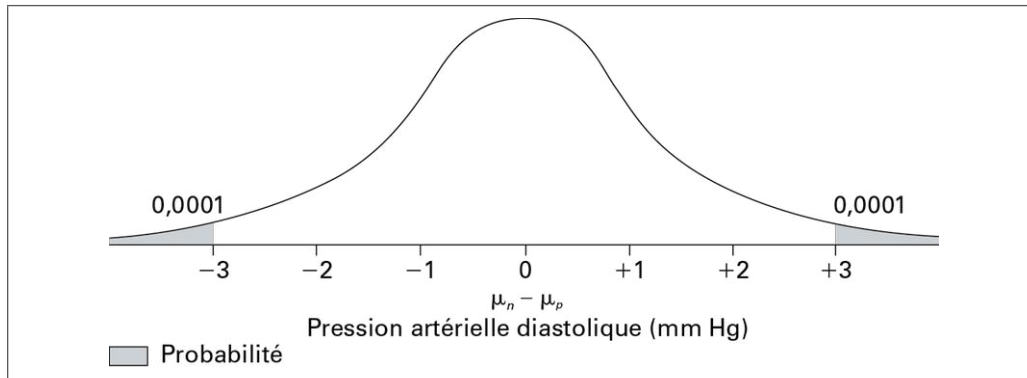
Dans l'étude, les chercheurs ont observé un écart de -3 mm Hg entre les deux moyennes de la PAD. La probabilité d'obtenir un tel résultat nous est donnée par le calcul du score Z , ce calcul et celui de la probabilité correspondante pour une différence de moyennes entre deux échantillons sont présentés dans l'annexe 11.1. Ici, cela veut dire que si H_0 est vraie dans la population des hommes hypertendus, la probabilité d'observer au moins cet écart est inférieure à $0,0001$. Pour faciliter certains calculs, on dira que cette probabilité est égale à $0,0001$.

Puisque H_1 est *a priori* une hypothèse bilatérale, on doit prendre en compte la possibilité que le Novel soit plus efficace que le Périn et celle qu'il soit moins efficace. On doit donc considérer également la probabilité d'observer un écart de $+3$ mm Hg et plus. Cette probabilité est aussi de $0,0001$. Alors, la probabilité d'observer un écart plus grand que -3 mm Hg ou que $+3$ mm Hg est de $0,0002$, les deux probabilités s'additionnant; cette observation peut être représentée par la figure 11.6.

Cette probabilité d'obtenir par hasard un résultat aussi extrême ou plus extrême que celui observé dans les échantillons, si H_0 est vraie, s'appelle la **valeur p**. Évidemment,

plus la valeur p s'avère petite, moins H_0 est jugée vraisemblable. La valeur p de 0,0002 étant faible, l'écart est peu compatible avec H_0 . Alors, H_0 pourrait être rejetée au profit de H_1 , jugée plus vraisemblable. On conclut alors que l'écart observé est, selon toute vraisemblance, non pas attribuable au hasard, mais bien à l'efficacité du nouveau médicament. On vient ainsi d'effectuer un test d'hypothèse.

FIGURE 11.6 ■ La probabilité d'observer un écart plus grand que - 3 mm Hg ou que + 3 mm Hg, si H_0 est vraie ($\mu_n = \mu_p$)



Avec une valeur p de 0,0002, il y a quand même un risque de se tromper 2 fois sur 10 000 si l'on rejette H_0 . L'erreur consistant à rejeter H_0 , alors qu'elle est vraie, correspond à l'erreur dite « de type I ».

À retenir

La valeur p est la probabilité d'obtenir par hasard un résultat aussi extrême ou plus extrême que celui observé dans l'échantillon si H_0 est vraie.

11.2.3 Le seuil de signification statistique (α)

Comment décider si le risque de se tromper, lorsqu'on rejette H_0 , est acceptable ou, en d'autres mots, comment décider si le risque de commettre une **erreur de type I** est acceptable? Le niveau de risque jugé acceptable de commettre une erreur de type I est fixé par le chercheur lui-même, par exemple selon la gravité des conséquences de faire ce type d'erreur. L'usage et la tradition suggèrent une probabilité de 0,05 ou de 0,01. Cette probabilité correspond au **seuil de signification statistique**, qu'on représente par la lettre grecque alpha (α). Le seuil de signification statistique correspond à la probabilité consentie de commettre une erreur de type I; il s'agit en quelque sorte d'un seuil de décision qui permet de rejeter ou non H_0 . Il est primordial de fixer cette probabilité avant le début de l'étude et de ne pas la modifier en cours de route.

11.2.4 L'interprétation du résultat du test d'hypothèse

Dans l'étude à laquelle Lisa s'intéresse, les chercheurs ont fixé le seuil de signification à 0,05. Cela veut dire qu'ils acceptent de se tromper 1 fois sur 20 lorsqu'ils

concluent, dans le traitement de l'HTA, à une différence d'efficacité entre le Novel et le Périn*. Si la valeur p calculée est égale ou inférieure au seuil de signification fixé, H_0 sera rejetée. Dans cette même étude, la valeur p calculée pour l'écart entre les moyennes de la PAD est égale à 0,0002. Comme cette valeur p s'avère plus petite que le seuil de signification fixé à 0,05, H_0 est rejetée, et les résultats obtenus à partir de l'échantillon sont alors dits « statistiquement significatifs pour le seuil de signification fixé ».

Ce dernier énoncé est la conclusion d'un test d'hypothèse. Il signifie que, tout en reconnaissant l'existence d'un certain risque d'erreur, on accepte que le nouveau médicament soit plus efficace que le médicament actuellement en usage.

Cependant, que faudrait-il conclure si la valeur p était de 0,06 ou de 0,25? Dans ces cas, le résultat ne serait pas considéré comme statistiquement significatif, puisque le seuil de signification a été fixé à 0,05. Ainsi, lorsqu'on établit ce seuil à 0,05, cela implique que toute valeur p supérieure à 0,05 conduit à la même conclusion, celle consistant à ne pas rejeter H_0 .

De façon générale, la valeur p est considérée comme une mesure de vraisemblance d'une hypothèse. Plus la valeur p est petite, plus grande est la vraisemblance de H_1 . Pour cette raison, il est suggéré, même si un résultat est jugé non statistiquement significatif, de présenter la valeur p et de laisser le lecteur l'interpréter en fonction de la taille des échantillons, de la nouveauté des résultats ou de l'intérêt qu'ils suscitent.

11.2.5 L'erreur de type II

Pour évaluer la rapidité d'action des deux médicaments, les chercheurs ont comparé la proportion des individus recevant le Novel et celle des individus recevant le Périn dont la PAD était redevenue normale (< 90 mm Hg) après deux mois de traitement. Ils ont observé une différence de proportion de 15 % en faveur du Novel. Cependant, même si cette différence semblait intéressante du point de vue clinique, ils ne l'ont pas jugée statistiquement significative, ce qui veut dire que la valeur p obtenue à l'aide d'un test statistique était plus élevée que le seuil fixé à 0,05. Doit-on conclure qu'il n'y a pas de différence de rapidité d'action entre les deux traitements dans la population à l'étude? Voilà certes une conclusion qui pourrait correspondre à la réalité. Toutefois, on peut aussi, à partir des résultats de l'échantillon, conclure à tort qu'il n'y a pas de différence de rapidité d'action entre les deux médicaments, alors que cette différence existe bel et bien dans la population; autrement dit, retenir H_0 alors que c'est H_1 qui est vraie. Cette erreur correspond à l'**erreur de type II**, et la probabilité de la commettre correspond à β (bêta).

Les conclusions qu'il serait possible de tirer à la suite de la confrontation des résultats d'un test d'hypothèse (jugement statistique) et de la vérité, même si, en pratique, celle-ci demeure inconnue, sont présentées dans le tableau 11.1. Quatre conclusions sont possibles. On peut rejeter ou non H_0 et avoir raison dans un cas comme dans l'autre puisque, dans le premier cas, c'est bien H_1 qui est vraie et que, dans le second,

* Sur le plan formel, selon la distribution de probabilité, cet énoncé signifie ceci: si l'on répétait un grand nombre de fois cette même expérience, H_0 serait rejetée 5 fois sur 100, alors que c'est pourtant elle qui est vraie.

c'est plutôt H_0 qui l'est. En revanche, on peut faire exactement l'inverse, c'est-à-dire rejeter H_0 alors que c'est elle qui est vraie ou la retenir, alors que c'est H_1 qui est vraie.

TABEAU 11.1 ■ Les conclusions possibles à la suite de la confrontation des résultats d'un test d'hypothèse (jugement statistique) et de la vérité

Jugement statistique	Vérité : H_1 vraie	Vérité : H_0 vraie
Rejet de H_0	Aucune erreur : jugement concordant avec la vérité	Erreur de type I : rejet de H_0 alors que H_0 est vraie
Non-rejet de H_0	Erreur de type II : non-rejet de H_0 alors que H_1 est vraie	Aucune erreur : jugement concordant avec la vérité

Si l'on revient à la question de la comparaison de la rapidité d'action des deux médicaments, on souhaite, bien sûr, éviter l'erreur qui consisterait à ne pas reconnaître une différence entre deux traitements (donc ne pas rejeter H_0) lorsque celle-ci existe réellement dans la population (c'est-à-dire lorsque H_1 est vraie). Toutefois, jusqu'à présent, on n'a pas précisé l'ampleur de cette différence entre les deux proportions. Voyons pourquoi il importe de le faire. Intuitivement, il est raisonnable de penser que, pour un seuil de signification fixé, plus la différence observée entre les proportions est grande, plus il devient facile de juger de la signification statistique des résultats obtenus avec un échantillon de petite taille. Par contre, plus cette différence observée dans la population s'avère petite, plus la taille de l'échantillon doit être grande pour permettre de conclure quant à la signification statistique des résultats. Cela signifie que la probabilité de commettre une erreur de type II est directement liée à cette différence dans la population. Dans l'étude à laquelle Lisa s'intéresse, les chercheurs ont observé une différence de 15 %. Si l'on répétait cette expérience plusieurs fois à partir d'autres échantillons de même taille, on obtiendrait différentes valeurs pour ces différences de proportions, par exemple 12 %, 16 % ou 26 %. Donc, à chacun de ces résultats, pour une même taille d'échantillon, correspondrait une probabilité différente de commettre une erreur de type II et de retenir à tort H_0 . Nous avons vu que le seuil de signification statistique (α) est fixé à l'avance. De même, la probabilité de commettre une erreur de type II, soit β , est une valeur pouvant être fixée à l'avance pour déterminer, par exemple, une taille d'échantillon ou une différence à observer. Cependant, contrairement au seuil de signification statistique, la probabilité de commettre une erreur de type II peut aussi se calculer à la fin de l'étude à partir d'éléments connus tels que la variabilité de la mesure, le seuil de signification fixé, la taille des échantillons et la différence observée.

À retenir

- L'erreur de type I est l'erreur commise si l'on rejette H_0 alors qu'elle est vraie, et l'erreur de type II est l'erreur commise si l'on ne la rejette pas alors qu'elle est fausse.
- Le seuil de signification statistique (α) est la probabilité consentie de commettre une erreur de type I, et bêta (β) est la probabilité consentie, ou calculée, de commettre une erreur de type II.

11.2.6 La puissance statistique

La valeur complémentaire de β , soit $1 - \beta$, s'appelle la **puissance de l'étude**. La puissance d'une étude est, pour une taille d'échantillon donnée, la probabilité de déceler une certaine différence lorsque celle-ci existe vraiment dans la population, donc la probabilité de rejeter H_0 alors que H_1 est vraie. Il est à noter que dans la littérature, les termes puissance, **puissance statistique** et puissance statistique d'une étude sont équivalents.

EXEMPLE 11.1

Dans l'étude qui intéresse Lisa, la puissance correspond à la capacité de mettre en évidence, à partir de 2 échantillons de 40 hommes hypertendus traités avec le Novel ou le Périn, une différence de proportion de 15 % lorsque cette différence existe vraiment dans la population à l'étude. Une différence de 15 % peut être considérée comme étant relativement petite, et, sur le plan statistique, ce nombre de 40 hommes par échantillon peut s'avérer insuffisant pour permettre de généraliser une telle différence avec confiance à la population cible. Il est donc possible que cette étude ne soit pas assez puissante pour conclure qu'une différence de 15 % est statistiquement significative pour un seuil de signification fixé à 0,05.

Plusieurs facteurs permettent d'améliorer la puissance d'une étude, et la taille des échantillons est le plus important d'entre eux.

Il est possible, avant le début d'une étude, de déterminer le nombre de sujets nécessaires pour déceler une certaine différence avec une certaine puissance⁶. En règle générale, une puissance de l'ordre de 80 % est jugée acceptable. C'est bien sûr l'intérêt clinique d'une certaine différence qui détermine en bonne partie la pertinence de vouloir la déceler.

EXEMPLE 11.2

Il serait peu intéressant, du point de vue clinique, de réaliser une étude afin de déceler une différence de survie moyenne de quelques jours chez des patients cancéreux soumis à un nouveau traitement antinéoplasique.

Lorsque les résultats d'une étude sont considérés comme statistiquement significatifs (donc lorsqu'on rejette H_0), il n'est pas pertinent de se questionner sur la puissance de l'étude, car le fait de rejeter H_0 implique nécessairement que la puissance de l'étude était suffisante pour détecter la différence observée. À l'inverse, lorsque les résultats d'une étude ne sont pas considérés comme statistiquement significatifs, on doit s'interroger sur sa puissance. Plus la puissance est élevée, plus les conclusions statistiques inspireront confiance. Par contre, si l'étude n'est pas assez puissante, la prudence sera de mise avant de conclure définitivement en ce qui a trait à ses résultats.

Donc, avant de considérer comme vrai le fait qu'il n'y a pas de différence entre la rapidité d'action du Novel et celle du Périn, il faudrait vérifier la puissance statistique de l'étude. Malheureusement, cette information n'est pas toujours présentée dans les publications scientifiques.

À retenir

La puissance statistique est la probabilité de détecter une certaine différence lorsqu'elle existe vraiment dans la population.

11.3 L'intervalle de confiance

Le test d'hypothèse ne procure aucune information sur la valeur des paramètres étudiés. Les moyennes des deux échantillons sont en fait des estimations ponctuelles du paramètre de chacune des populations cibles. Ces estimations sont sujettes aux fluctuations d'échantillonnage, et la probabilité qu'une telle estimation ponctuelle soit égale au paramètre est presque nulle. Il existe un moyen de mieux estimer ce paramètre dans chacune des deux populations à l'étude : l'**intervalle de confiance**, qui consiste à trouver un intervalle de valeurs qui a de bonnes probabilités de contenir la valeur du paramètre de la population cible.

EXEMPLE 11.3

Si l'on essaie de déterminer le poids exact d'une personne, on a peu de chances d'y arriver. Cependant, si l'on tente de l'estimer avec un intervalle (± 5 kg), les chances que son poids soit compris dans cet intervalle sont plus grandes, et elles le deviennent encore davantage si l'on essaie de l'estimer à ± 10 kg. En effet, supposons que le poids réel de celle-ci soit de 62 kg et qu'on l'estime à 65 kg avec un intervalle de ± 5 kg (soit de 60 à 70 kg); les chances sont plus grandes que son poids réel soit compris dans l'intervalle que celles d'avoir misé juste avec l'estimation de 65 kg, et elles seront encore plus grandes pour un intervalle allant de 55 à 75 kg (65 ± 10 kg). Cependant, dans le cas de ce dernier intervalle, la précision de l'estimation s'avère plus faible.

Dans l'étude dont il est actuellement question, pour mieux estimer les paramètres de la moyenne de la PAD des deux populations d'hommes hypertendus, soit celle traitée au Novel et l'autre recevant le Périn, on pourrait présenter une estimation par intervalle à partir des résultats obtenus dans les échantillons. Cependant, cet intervalle doit avoir une probabilité assez élevée de recouvrir le paramètre. Cette probabilité correspond au niveau de confiance de l'intervalle. On peut penser que plus l'intervalle est large, plus la probabilité de recouper le paramètre est forte. Par contre, un intervalle trop large conduit à une estimation peu précise du paramètre. Il faut trouver un juste milieu entre le niveau de confiance désiré et la précision de l'estimation, soit la largeur de l'intervalle.

Dans l'étude, l'estimation ponctuelle de la moyenne de la PAD pour le Novel est de 86 mm Hg. Si l'on présente une estimation par intervalle de 86 ± 3 mm Hg, c'est-à-dire de 83 mm Hg à 89 mm Hg, on a certes beaucoup de chances de recouper le paramètre; en revanche, l'intervalle sera large, ce qui entraîne un manque de précision de l'estimation. Par ailleurs, un intervalle de 86 ± 1 mm Hg, c'est-à-dire

de 85 mm Hg à 87 mm Hg, est moins large et donc plus précis; mais quelle est la probabilité qu'il recouvre le paramètre? En d'autres termes, quelle confiance a-t-on que cet intervalle recouvre le paramètre et comment déterminer le niveau de confiance acceptable?

Tout comme le seuil de signification d'un test d'hypothèse, le niveau de confiance d'un intervalle est fixé par le chercheur. Il s'agit d'un choix arbitraire, mais on recommande de conserver une probabilité de recouvrir la valeur du paramètre de 95%. On parle alors d'un «intervalle de confiance à 95%». Toutefois, rien n'empêche le chercheur de choisir un niveau de confiance à 90% ou à 99%. Le calcul d'un intervalle de confiance pour une moyenne est présenté dans l'annexe 11.3.

EXEMPLE 11.4

Supposons que dans l'échantillon des hommes hypertendus traités avec le Novel, les limites inférieure et supérieure de l'intervalle de confiance de la moyenne de la PAD sont de 85 mm Hg et 87 mm Hg. Il y a donc une probabilité de 95% que l'intervalle de confiance qui s'étend de 85 mm Hg à 87 mm Hg recouvre la vraie valeur de la moyenne de la PAD, c'est-à-dire le paramètre, observée dans la population des hommes hypertendus traités avec le Novel. Pour le paramètre de la population des individus hypertendus traités avec le Périn, l'intervalle de confiance à 95% s'étend de 88 mm Hg à 90 mm Hg.

À retenir

L'intervalle de confiance permet de juger de la précision d'une estimation ponctuelle faite à partir d'un échantillon de la population. En fait, pour un même niveau de confiance, disons 95%, plus l'intervalle sera étroit et plus l'estimation sera précise.

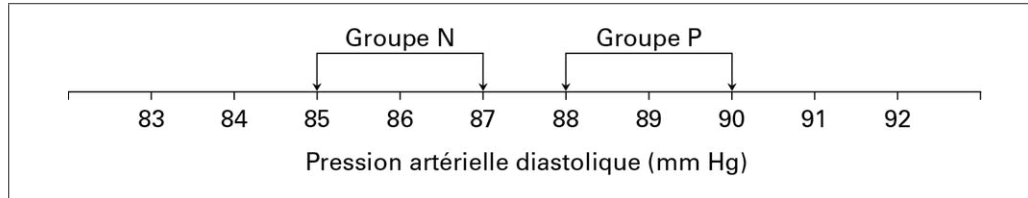
11.3.1 La vérification d'une hypothèse à l'aide d'un intervalle de confiance

La comparaison d'intervalles de confiance

Nous avons vu qu'on peut juger de la signification statistique lorsqu'on compare deux moyennes en ayant recours à un test statistique approprié, en calculant une valeur p et en vérifiant si elle est égale ou inférieure au seuil de signification fixé. On peut aussi comparer les intervalles de confiance des deux moyennes, ce qui constitue cependant une méthode plus approximative pour réaliser un test d'hypothèse. Il y a donc deux façons de juger de la signification statistique lorsqu'on compare deux moyennes.

En comparant les intervalles de confiance à 95% des moyennes de la PAD des deux groupes au moyen d'une figure (voir la figure 11.7), on constate que les intervalles de confiance ne se recoupent pas. On peut alors conclure que la différence observée entre les moyennes de la PAD des deux groupes est statistiquement significative pour un seuil de signification fixé à 0,05.

FIGURE 11.7 ■ L'intervalle de confiance à 95 % pour les valeurs de la moyenne de la PAD (mm Hg) des deux groupes de traitement (groupe N et groupe P)



Dans le cas où les deux intervalles à 95 % ne se recoupent pas, la différence entre les deux moyennes est jugée statistiquement significative pour un seuil de signification fixé à 0,05. Cependant, s'il y a un recouvrement des deux intervalles de confiance, on ne pourra pas tirer de conclusion.

EXEMPLE 11.5

Il y aurait recouvrement entre les deux intervalles de confiance comparés si, par exemple, l'intervalle de confiance à 95 % de la moyenne de la PAD pour le Novel allait de 84 mm Hg à 88 mm Hg et, pour le Périn, de 87 mm Hg à 91 mm Hg. Les valeurs comprises entre 87 et 88 sont alors présentes dans les deux intervalles comparés.

Deux approches peuvent dans ce cas nous permettre d'arriver à une conclusion :

- 1) le recours à un test statistique de comparaison de deux moyennes (ou deux proportions) ;
- 2) la construction d'un intervalle de confiance autour de la différence de moyennes (ou de proportions).

Dans la première approche, on interprète la valeur p obtenue. Dans la seconde approche, si la valeur nulle (0) se retrouve à l'intérieur de l'intervalle de confiance de la différence de moyennes, on ne rejettera pas H_0 ; à l'inverse, si 0 ne se trouve pas dans cet intervalle, on rejettera H_0 . En comparant le Novel et le Périn, on a obtenu une différence de moyenne de -3 mm Hg. Si l'on construit un intervalle de confiance à 95 % autour de cette différence et que l'on obtient des bornes de -5 et -1 par exemple, on conclura alors qu'il y a une différence entre les deux groupes (en acceptant une erreur de type I de 5 %). En effet, la valeur 0 n'est pas comprise dans cet intervalle.

L'intervalle de confiance et les mesures d'association

On peut également utiliser l'intervalle de confiance comme test d'hypothèse pour d'autres mesures d'association utilisées en épidémiologie, comme un risque relatif (RR) ou un rapport de cotes (RC). Pour ces mesures, l'absence d'association entre deux variables se traduit par une valeur de 1 puisqu'on parle de rapports. Ainsi, en vérifiant si la valeur 1 est comprise ou non dans l'intervalle de confiance calculé autour d'une telle mesure d'association, il est possible de conclure quant à la signification statistique de l'association entre les deux variables.

À partir d'un échantillon, lorsqu'on étudie une association entre une exposition et une maladie à l'aide des mesures d'incidences cumulées, on a le choix entre deux stratégies pour juger de la signification statistique des résultats. Premièrement, on peut comparer la différence entre les deux incidences cumulées, c'est-à-dire la proportion de malades chez les exposés et la proportion de malades chez les non-exposés, à l'aide d'un test statistique approprié, pour déterminer ensuite une valeur p et finalement vérifier si elle est égale ou inférieure au seuil de signification fixé à 0,05. Deuxièmement, on peut calculer la mesure d'association correspondante, soit le RR, et calculer l'intervalle de confiance à 95 % pour cette mesure. Si l'intervalle de confiance à 95 % ne recouvre pas la valeur 1, c'est-à-dire la valeur de non-association, le RR calculé à partir de l'échantillon est jugé statistiquement significatif pour un seuil de signification fixé à 0,05. Par contre, si l'intervalle recouvre la valeur 1, le résultat ne sera pas statistiquement significatif.

EXEMPLE 11.6

Si l'on étudie la relation entre la consommation de cigarettes et les maladies cardiaques dans un échantillon de la population, et que le RR calculé est de 2,1 avec un intervalle de confiance à 95 % qui s'étend de 1,3 à 3,5, on conclura que le RR est statistiquement significatif pour un seuil de signification fixé à 0,05.

Pour chaque type de mesure en épidémiologie, il existe une équation particulière servant à calculer l'intervalle de confiance désiré. L'important ici n'est pas de savoir comment calculer ces intervalles, mais d'en faire une interprétation correcte.

EXEMPLE 11.7

Si l'auteur d'une étude cas-témoins calcule un rapport de cotes, soit un RC, égal à 1,8, avec un intervalle de confiance à 95 % qui s'étend de 0,7 à 2,9, alors ce RC ne sera pas jugé comme statistiquement significatif pour un seuil de signification fixé à 0,05, et l'on pourra se questionner sur la puissance statistique de l'étude.

À retenir

L'intervalle de confiance est un intervalle de valeurs qui a une certaine probabilité de contenir la valeur du paramètre.

11.4 La signification statistique et la signification clinique

Lisa devrait-elle prescrire le nouveau médicament à Charles? Dans sa prise de décision, après avoir porté un jugement sur la validité et la signification statistique des résultats de l'étude, elle doit absolument tenir compte de leur **signification clinique**.

Une différence significative sur le plan statistique ne l'est pas nécessairement sur le plan clinique. Dans l'étude portant sur le Novel, après deux mois de traitement, les moyennes de la PAD des groupes N et P sont respectivement de 86 mm Hg et de 89 mm Hg. L'écart observé est différent de 0 et statistiquement significatif pour un seuil de signification fixé à 0,05. Cependant, le nouveau médicament présente-t-il pour autant un intérêt clinique ?

Pour répondre à cette question, il faut d'abord juger de l'importance clinique de l'écart. Il revient au clinicien de déterminer, à partir des connaissances qu'il possède sur le sujet, si un écart de -3 mm Hg présente un quelconque intérêt clinique. Il s'agit donc ici d'un jugement clinique et non statistique. La valeur p ne permet pas de porter un jugement sur l'importance clinique des résultats.

Outre l'importance clinique des résultats, il faut tenir compte de certains autres facteurs.

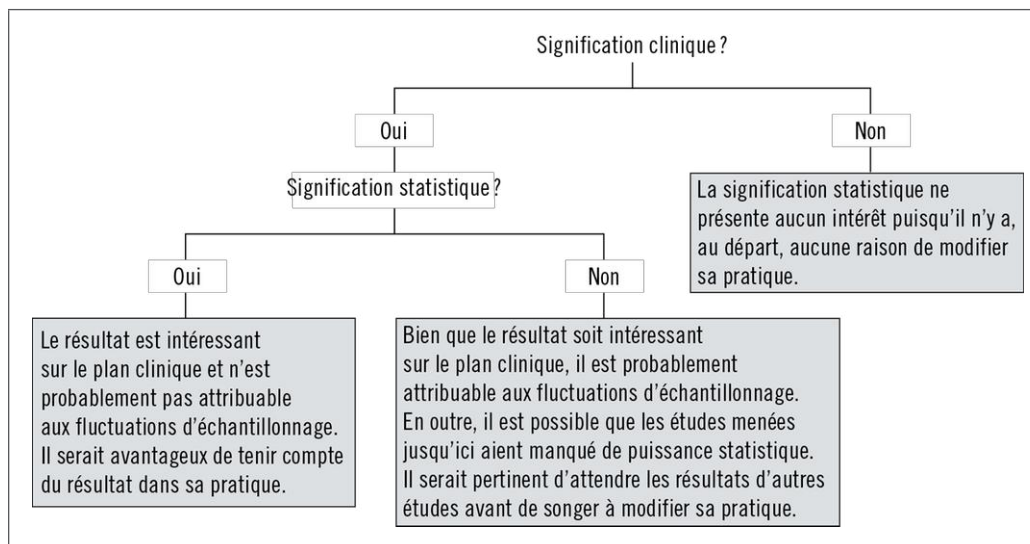
- Le profil d'effets secondaires du nouveau médicament est-il connu ?
- Ce produit cause-t-il moins d'effets secondaires que le médicament en usage actuellement ?
- Quelles seront les répercussions du coût et des modalités posologiques du nouveau médicament sur l'adhésion au traitement ?
- La très courte période d'observation (deux mois) permet-elle d'évaluer l'innocuité à long terme du nouveau médicament ?

De la même façon, la prudence s'impose avant de conclure qu'un résultat qui n'est pas statistiquement significatif ne présente aucun intérêt sur le plan clinique. C'est le cas de la différence de rapidité d'action entre le Novel et le Périn, pour lesquels, au terme de deux mois de traitement, respectivement 80 % et 65 % des individus ont une PAD normale, soit une différence de 15 %. Bien qu'elle ne soit pas significative sur le plan statistique, cette différence peut sembler intéressante d'un point de vue clinique. Dans un tel cas, il est très important de vérifier la puissance de l'étude. Si elle est insuffisante, il faudra attendre les résultats d'autres études avant de porter un jugement définitif.

Finalement, la signification clinique correspond à une différence jugée importante et qui porte à conséquence sur le plan clinique. La **signification statistique** fait plutôt référence à la probabilité que la différence observée dans les données soit attribuable au hasard (aux fluctuations d'échantillonnage). La figure 11.8, à la page suivante, présente, sous la forme d'un algorithme, la démarche qui permet, en fonction des significations clinique et statistique d'un résultat, de juger de la pertinence de modifier ou non sa pratique.

Après une lecture attentive de la publication portant sur l'utilité du Novel dans le traitement de l'HTA, il s'avère pertinent d'attendre les résultats d'autres études sur le sujet avant de le prescrire. En effet, l'efficacité et l'innocuité à long terme du Novel n'ont pas été établies, les auteurs n'ont pas précisé la puissance de l'étude, et ils n'ont pas fait mention des effets secondaires potentiels de ce nouveau médicament. En revanche, le Périn, médicamenteusement efficace et dont l'innocuité a été établie, a fait l'objet de plusieurs publications.

FIGURE 11.8 ■ La démarche permettant, en fonction de la signification clinique et de la signification statistique d'un résultat, de juger de la pertinence de modifier ou non sa pratique



En résumé

L'inférence statistique est un processus qui permet d'induire la valeur d'un paramètre à partir d'une mesure obtenue dans un échantillon représentatif de la population. Le test d'hypothèse et l'intervalle de confiance sont deux approches statistiques qui relèvent de l'inférence statistique. Le test d'hypothèse est une démarche qui permet de faire, sur le plan statistique, le choix le plus judicieux entre deux hypothèses statistiques: l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1). Le seuil de signification statistique (α) d'un test d'hypothèse est la probabilité consentie de rejeter H_0 quand elle est vraie. À l'aide d'un test statistique, on détermine une valeur p à partir des résultats de l'échantillon. La valeur p est la probabilité d'obtenir par hasard

un résultat aussi extrême ou plus extrême que celui observé dans l'échantillon si H_0 est vraie. L'obtention d'une petite valeur p (inférieure au seuil α fixé) permet de conclure quant à la signification statistique d'un résultat, mais elle ne compense en rien les faiblesses d'une étude mal construite et comportant des biais importants.

L'intervalle de confiance permet d'apprécier la précision d'une estimation faite à partir d'un échantillon.

Un résultat significatif sur le plan statistique ne l'est pas nécessairement sur le plan clinique. Sur le plan clinique, les résultats doivent être jugés utiles et tenir compte des répercussions sur la santé des individus.

Exercices



Voir la section complète des exercices sur le Web.

1. Lequel de ces outils statistiques devez-vous utiliser pour inférer un taux d'incidence observé dans un échantillon à une population cible?
 - a) Le test d'hypothèse.
 - b) L'intervalle de confiance.
2. Vrai ou faux? Plus la valeur p est faible, moins l'hypothèse nulle est considérée comme vraisemblable et plus on sera porté à rejeter H_0 .
3. Lequel de ces énoncés est vrai?
 - a) L'erreur de type I est une erreur que l'on fait lorsqu'on rejette H_0 , alors qu'en réalité c'est elle qui est vraie.
 - b) L'erreur de type I est une erreur que l'on fait lorsqu'on accepte H_0 , alors qu'en réalité elle est fausse.
4. Si nous obtenons le résultat $RR = 16$, $p < 0,05$, alors :
 - a) nous devons rejeter H_0 avec un seuil de signification à 0,05.
 - b) nous devons accepter H_0 avec un seuil de signification à 0,05.
5. Parmi les énoncés suivants, lequel des intervalles de confiance à 95 % est le plus précis?
 - a) $RR = 16$, intervalle de confiance à 95 % (14; 18).
 - b) $RR = 16$, intervalle de confiance à 95 % (13; 19).

Pour les réponses, voir le corrigé à la fin du manuel.

ANNEXE

11.1

Test d'hypothèse pour la comparaison de deux moyennes

Lorsqu'on compare deux moyennes, deux situations peuvent se présenter. Celles-ci se définissent par deux distributions d'échantillonnage différentes déterminées principalement par la taille des échantillons.

La première situation s'apparente à l'exemple de l'étude visant à évaluer les mérites du nouveau médicament antihypertenseur, le Novel, discutée dans le présent chapitre, dans laquelle les échantillons présentent les caractéristiques suivantes.

- Les populations cibles sont importantes (population de tous les hommes hypertendus).
- Les deux échantillons sont formés de façon aléatoire.
- La taille respective des deux échantillons est de 30 individus ou plus.

Dans ce cas, la démarche du test d'hypothèse se résume comme suit. On veut comparer la moyenne de la PAD de deux populations d'hommes hypertendus recevant deux médicaments différents (le Novel et le Périn). La comparaison est basée sur l'écart entre les moyennes de la PAD observées dans les deux échantillons. La distribution d'échantillonnage de l'écart entre les deux moyennes est connue; il s'agit d'une

distribution normale. Il est donc possible de calculer un score Z et de déterminer sa valeur p . Cette valeur p est comparée au seuil de signification fixé afin de juger de la signification statistique des résultats.

Le test d'hypothèse peut être réalisé en cinq étapes.

1. Formuler les hypothèses statistiques en tenant compte des deux populations.
 - Hypothèse nulle (H_0): la moyenne de la PAD dans la population des hommes hypertendus traités avec le Novel (groupe N) = la moyenne de la PAD dans la population des hommes hypertendus traités avec le Périn (groupe P).
 - Hypothèse alternative (H_1): la moyenne de la PAD dans la population des hommes hypertendus traités avec le Novel \neq la moyenne de la PAD dans la population des hommes hypertendus traités avec le Périn.
2. Fixer le seuil de signification statistique.
 - Les chercheurs ont fixé ce seuil à 0,05.
3. Choisir le test statistique en relation avec la loi normale centrée et réduite pour la comparaison de deux moyennes.
 - La formule du score Z servant au calcul d'une différence de moyennes est la suivante:

$$Z = \frac{\bar{x}_n - \bar{x}_p}{\sqrt{\frac{s_n^2}{n_n} + \frac{s_p^2}{n_p}}}$$

4. Calculer le score Z (valeur de l'abscisse dans la table de la distribution normale centrée et réduite).
 - On pose que:
 - \bar{x}_n = moyenne de la PAD du groupe N = 86 mm Hg;
 - \bar{x}_p = moyenne de la PAD du groupe P = 89 mm Hg;
 - $n_n = n_p = 40$ (taille des échantillons);
 - $s_n = s_p = 3,2$ mm Hg (écart type).
 - Le score Z est donc de 4,19.
5. Juger de la signification statistique des résultats.
 - En consultant la table de la distribution normale centrée et réduite (*voir le tableau 10.4, p. 211-212*), on constate, si H_0 est vraie, qu'un score Z de $-4,19$ et moins (ou de $4,19$ et plus) correspond à une valeur $p < 0,0001$. Cette valeur p est inférieure au seuil de signification fixé. Donc H_0 est rejetée, et l'on accepte H_1 . La différence entre les moyennes de la PAD (-3 mm Hg) est jugée statistiquement significative pour un seuil de signification fixé à 0,05.

La deuxième situation survient lorsque la taille d'au moins un des deux échantillons est inférieure à 30 individus. Dans ce cas, on doit supposer que les échantillons proviennent de populations où la distribution des écarts entre les moyennes est normale et que les variances sont identiques. Alors, les fluctuations d'échantillonnage se

distribuent selon la loi de Student. À partir de la moyenne et de l'erreur type de cette distribution, il est possible de construire un test statistique, le **test de t** , qui permet, à l'aide d'une table de probabilités, la table de la distribution de Student, de déterminer une valeur p .

ANNEXE 11.2 Test d'hypothèse pour la comparaison de deux proportions

Comme pour la comparaison de deux moyennes, on doit connaître la distribution d'échantillonnage de la différence de deux proportions si H_0 est vraie. En règle générale, les fluctuations d'échantillonnage de la différence de deux proportions se répartissent selon la loi normale centrée et réduite si les échantillons sont tirés au hasard et de façon indépendante, et si la taille des deux échantillons est égale ou supérieure à 30 individus. Si ces conditions sont respectées, le test d'hypothèse peut être réalisé en cinq étapes.

1. Formuler les hypothèses statistiques en tenant compte des deux populations.
 - Hypothèse nulle (H_0): le pourcentage d'hommes normotendus après deux mois de traitement dans la population d'hommes hypertendus recevant le Novel (groupe N) = le pourcentage d'hommes normotendus après deux mois de traitement dans la population d'hommes hypertendus recevant le Périn (groupe P).
 - Hypothèse alternative (H_1): le pourcentage d'hommes normotendus après deux mois de traitement dans la population d'hommes hypertendus recevant le Novel \neq le pourcentage d'hommes normotendus après deux mois de traitement dans la population d'hommes hypertendus recevant le Périn.
2. Fixer le seuil de signification statistique.
 - Les chercheurs ont fixé ce seuil à 0,05.
3. Choisir le test statistique en relation avec la loi normale centrée et réduite pour la différence de deux proportions.
 - La formule du score Z servant au calcul d'une différence de proportions est la suivante:

$$Z = \frac{p_n - p_p}{\sqrt{\frac{p_n q_n}{n_n} + \frac{p_p q_p}{n_p}}}$$

4. Calculer le score Z (valeur de l'abscisse dans la table de la distribution normale centrée et réduite).

- On pose que :
 - p_n = proportion d'hommes normotendus traités avec le Novel = 0,80, alors $q_n = 0,20$;
 - p_p = proportion d'hommes normotendus traités avec le Périn = 0,65, alors $q_p = 0,35$;
 - $n_n = n_p = 40$ (taille de l'échantillon).
 - Le score Z est donc de 1,52.
5. Juger de la signification statistique des résultats.

- En consultant la table de la distribution normale centrée et réduite (voir le tableau 10.4, p. 211-212), on constate, si H_0 est vraie, qu'un score Z de 1,52 et plus correspond à une valeur $p > 0,05$, le seuil de signification fixé. On ne peut pas rejeter H_0 . La différence entre les pourcentages d'hommes normotendus (15 %) n'est pas jugée statistiquement significative pour un seuil de signification fixé à 0,05.

Il importe de noter que, selon les conditions liées à la distribution d'échantillonnage et à la taille des échantillons, il existe d'autres tests statistiques permettant de comparer des proportions; entre autres, le **test du khi carré** (χ^2) permet de comparer deux proportions ou plus entre elles.

ANNEXE 11.3 Intervalle de confiance d'une moyenne

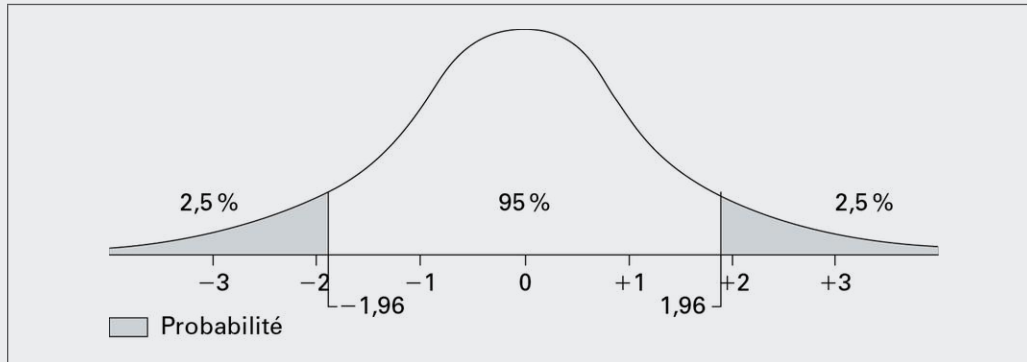
Grâce au théorème de la limite centrale, on peut calculer l'intervalle de confiance d'une moyenne. On doit cependant disposer de trois informations: la distribution des probabilités de la mesure, c'est-à-dire la moyenne en ce qui concerne l'étude; le niveau de confiance désiré, qui s'exprime par $1 - \alpha$; et l'erreur type de la mesure obtenue dans l'échantillon. On utilise ensuite une formule comme celle qui suit.

$$\text{moyenne} \pm \left[Z_{1-\alpha} \times \frac{s}{\sqrt{n}} \right]$$

Pour le Novel, le théorème de la limite centrale énonce que la distribution des moyennes échantillonnales de la PAD suit une distribution normale. Ce théorème permet d'utiliser la table de la distribution normale centrée et réduite afin de déterminer le score Z correspondant au niveau de confiance désiré. Un niveau de confiance établi à 95 %, c'est-à-dire exprimé par $1 - 0,05$, signifie que l'intervalle doit contenir 95 % des événements possibles. Les autres 5 % d'événements possibles doivent se répartir également, soit par tranches de 2,5 %, de chaque côté de l'intervalle de confiance à 95 %. La table de la distribution normale centrée et réduite (voir le tableau 10.4)

indique que, pour obtenir une probabilité de 0,025 dans chacune des extrémités de la courbe normale, le score Z doit être de 1,96, ce qui est représenté par la figure 11.9.

FIGURE 11.9 ■ L'intervalle de confiance à 95 % par rapport à la distribution normale centrée et réduite



Il ne reste plus maintenant qu'à calculer l'erreur type. Dans l'article, les chercheurs précisent que l'écart type des échantillons pour les deux groupes de traitement est le même, soit 3,2 mm Hg. Or, on connaît la taille des échantillons. L'erreur type sera donc de $3,2/\sqrt{40}$. L'intervalle de confiance à 95 % se calcule alors ainsi :

$$86 \pm \left[1,96 \times \frac{3,2}{\sqrt{40}} \right]$$

Le calcul conduit aux valeurs s'échelonnant de 85 mm Hg à 87 mm Hg. Ces valeurs constituent les limites inférieure et supérieure de l'intervalle de confiance. Il y a donc une probabilité de 95 % que l'intervalle qui s'étend de 85 mm Hg à 87 mm Hg recouvre la vraie valeur de la moyenne de la PAD, c'est-à-dire le paramètre, observée dans la population des hommes hypertendus traités avec le Novel.

Si l'on utilise la même formule pour obtenir une estimation par intervalle pour le paramètre de la population des hommes hypertendus traités avec le Périn, l'intervalle de confiance à 95 % s'étendra de 88 mm Hg à 90 mm Hg.

Collaboration à la 3^e édition :

Jocelyne Moisan

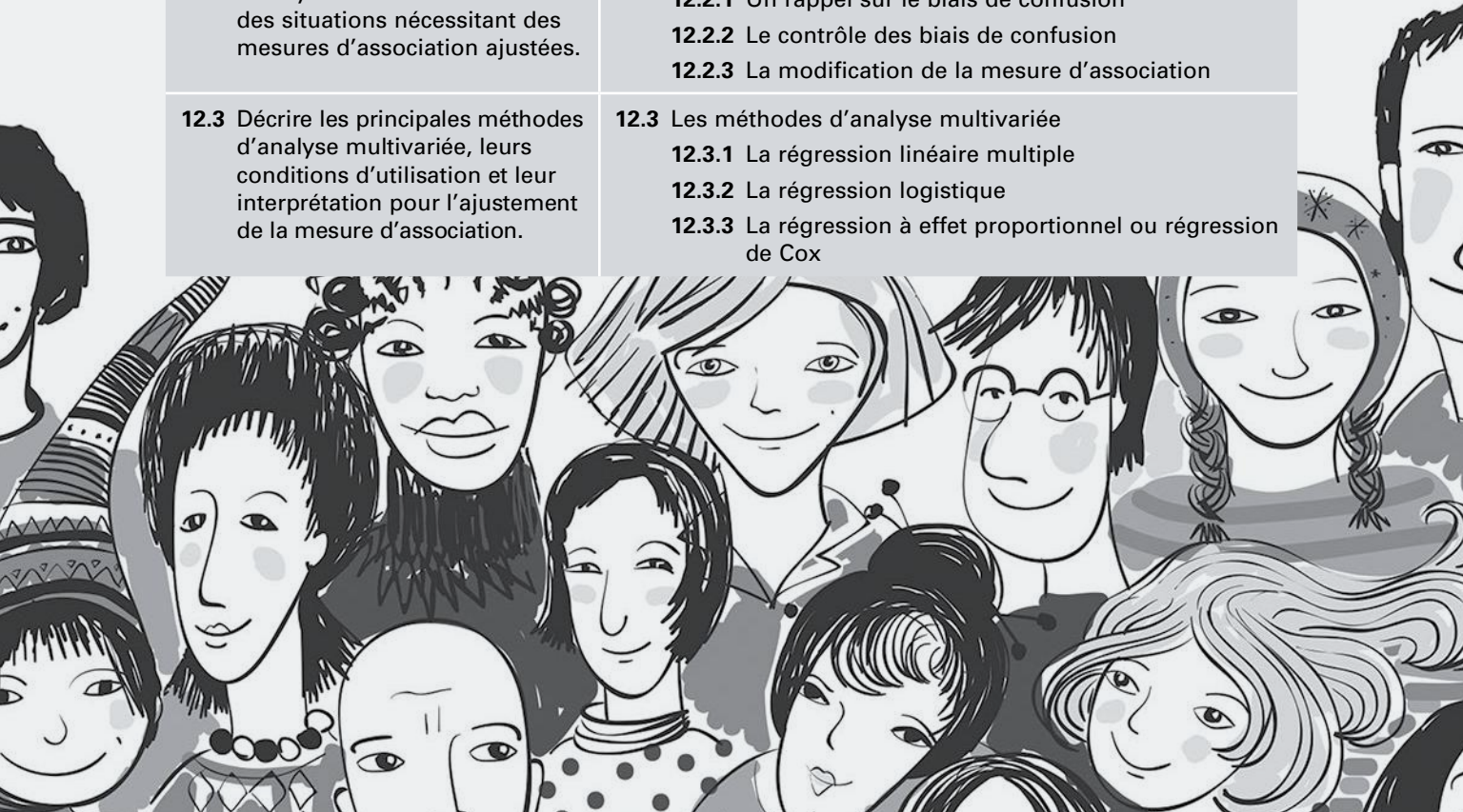
Jean-Pierre Grégoire

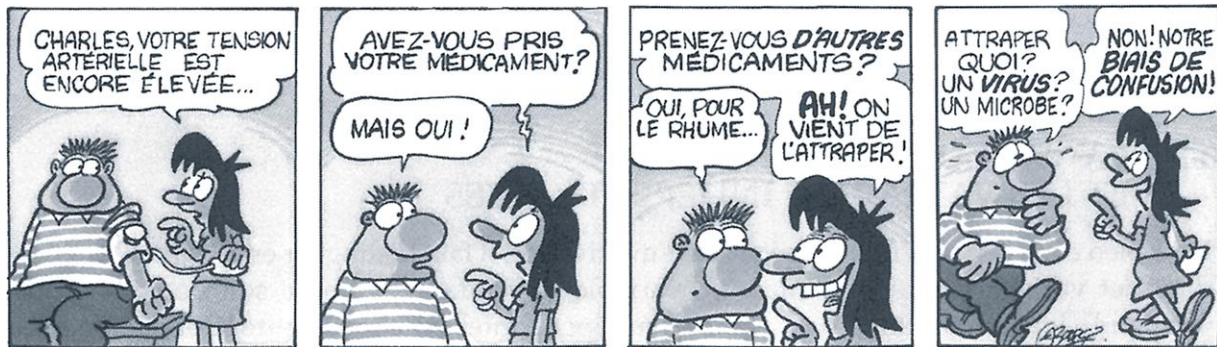
Suzanne Gingras

Au terme de ce chapitre, vous pourrez :

- expliquer les concepts et les méthodes de base de l'analyse multivariée;
- juger de la validité scientifique d'une étude et d'un résultat;
- prendre une décision éclairée sur la pertinence de modifier votre pratique clinique à la lumière des résultats d'études scientifiques.

Objectifs spécifiques	Contenu
12.1 Distinguer l'analyse univariée de l'analyse multivariée.	12.1 L'utilité de l'analyse univariée 12.1.1 La régression linéaire
12.2 Expliquer les fonctions de l'analyse multivariée et discuter des situations nécessitant des mesures d'association ajustées.	12.2 L'utilité de l'analyse multivariée 12.2.1 Un rappel sur le biais de confusion 12.2.2 Le contrôle des biais de confusion 12.2.3 La modification de la mesure d'association
12.3 Décrire les principales méthodes d'analyse multivariée, leurs conditions d'utilisation et leur interprétation pour l'ajustement de la mesure d'association.	12.3 Les méthodes d'analyse multivariée 12.3.1 La régression linéaire multiple 12.3.2 La régression logistique 12.3.3 La régression à effet proportionnel ou régression de Cox





Mise en situation

Voilà un mois, Lisa, omnipraticienne à la polyclinique Milo, a prescrit un médicament antihypertenseur à Charles. Elle mesure aujourd'hui sa pression artérielle et constate que celle-ci est encore anormalement élevée. Elle se demande pourquoi. Il se peut que la dose de médicament prescrite ne soit pas suffisante, que Charles ait omis de prendre son médicament ou qu'il n'ait pas pris toutes les doses prescrites. Il est aussi possible que Charles ait pris de façon concomitante un autre médicament dont l'un des effets aurait été d'augmenter la pression artérielle, qu'il ait consommé plus de sel, pris du poids ou diminué son niveau d'activité physique, tous ces facteurs étant reconnus comme susceptibles de modifier la pression artérielle. Lisa vérifie auprès de Charles si l'un de ces facteurs pourrait expliquer l'échec du traitement. Elle apprend que, depuis 10 jours, il prend un médicament contre le rhume. Lisa s'informe auprès de Marie-Louise pour savoir si ce médicament peut augmenter la pression artérielle, ce que lui confirme sa collègue pharmacienne. Cette dernière lui suggère de dire à son patient d'utiliser des vaporisations nasales de solution saline et de cesser de prendre son médicament contre le rhume. Lisa donne suite aux suggestions de Marie-Louise et prescrit à Charles

le même médicament antihypertenseur pour une période de un mois.

Pertinence clinique

Lorsque Lisa a constaté que la pression artérielle de Charles était toujours anormalement élevée après un mois de traitement, elle a envisagé plusieurs explications possibles. Elle a alors identifié des études qui ont recours à l'analyse multivariée. Ce type d'analyse permet de mesurer la part respective des différents facteurs qui peuvent expliquer l'échec d'un traitement, comme c'est ici le cas. La littérature scientifique abonde en études qui considèrent les effets de plusieurs facteurs sur le risque qu'une maladie donnée se développe. Des méthodes d'analyse multivariée sont habituellement utilisées pour prendre en compte l'effet simultané de plusieurs variables sur une autre variable. Le présent chapitre porte sur les principes de base de ces méthodes, sur leurs conditions d'application et sur l'interprétation que l'on doit faire des résultats obtenus. L'analyse multivariée ne sera pas abordée dans le détail; les lecteurs qui désirent approfondir leurs connaissances à ce sujet peuvent consulter des ouvrages spécialisés¹.

12.1 L'utilité de l'analyse univariée

Pour bien comprendre l'utilité de l'analyse multivariée, voyons tout d'abord la situation plus simple de l'analyse de la relation entre deux variables. En premier lieu, il faut établir une distinction entre la variable dépendante et la variable indépendante.

La **variable dépendante**, comme son nom l'indique, « dépend » en quelque sorte de la **variable indépendante**. Elle représente le phénomène ou l'événement à décrire. La variable indépendante représente pour sa part le phénomène ou l'événement qui peut influencer la variable dépendante. Ainsi, si l'on désire évaluer l'efficacité d'un traitement antihypertenseur, la variable indépendante est le traitement, alors que la variable dépendante est la pression artérielle.

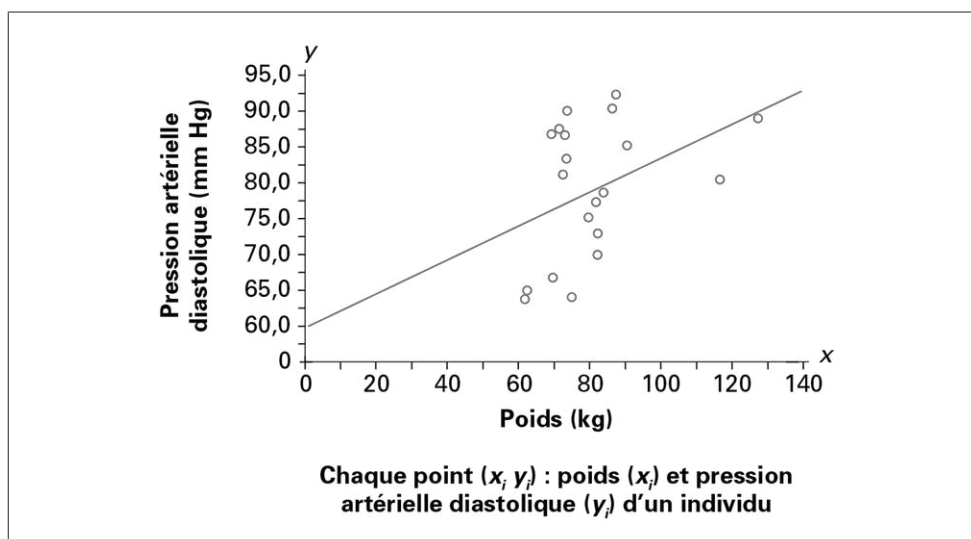
12.1.1 La régression linéaire

Plusieurs méthodes statistiques permettent d'établir la relation entre une variable dépendante et une variable indépendante. On utilise la **régression linéaire** lorsque la variable dépendante est une variable quantitative continue. Cette méthode produit un coefficient qui donne l'amplitude du changement de la variable dépendante pour chaque unité de la variable indépendante. Ainsi, elle rend compte de la relation qui existe entre le poids et la pression artérielle en indiquant que la variation du poids entraîne une variation de la pression artérielle.

EXEMPLE 12.1

Une étude menée auprès de 20 individus atteints d'hypertension artérielle (HTA) établit l'association entre la pression artérielle diastolique (la variable dépendante) et le poids (la variable indépendante). L'étude a été menée aux États-Unis auprès d'hommes blancs d'âge moyen. Les valeurs de la pression artérielle diastolique et les valeurs du poids obtenues pour chaque sujet sont inscrites dans la figure 12.1, laquelle représente d'ailleurs la relation entre ces deux variables.

FIGURE 12.1 ■ La relation entre la pression artérielle diastolique et le poids chez 20 États-Uniens blancs d'âge moyen



Le nuage de points qu'on y observe illustre le fait que plus les sujets sont lourds, plus leur pression artérielle diastolique est élevée.

La régression linéaire décrit donc la relation entre une variable indépendante (x_1) et une variable dépendante (y) en déterminant, par une opération mathématique, la droite qui s'ajuste le mieux au nuage de points, c'est-à-dire la droite qui décrit le mieux la relation entre les valeurs de ces deux variables x_1 et y . Les caractéristiques de cette droite sont représentées par le modèle mathématique suivant.

Régression linéaire (modèle mathématique 12.1): $y = \beta_0 + \beta_1 x_1$

Dans ce modèle, le symbole β_0 correspond à la valeur de y lorsque x_1 est égal à 0; il s'agit donc du point de rencontre entre la droite et l'axe des y . Ce point de rencontre β_0 est appelé l'«intersection» ou l'«ordonnée» à l'origine de y . Quant au symbole β_1 , il correspond à la pente de la droite de régression. Il indique dans quelle mesure un changement dans la valeur de x_1 influe sur la valeur de y .

Le modèle mathématique de la droite de régression correspondant à la figure 12.1 est le suivant.

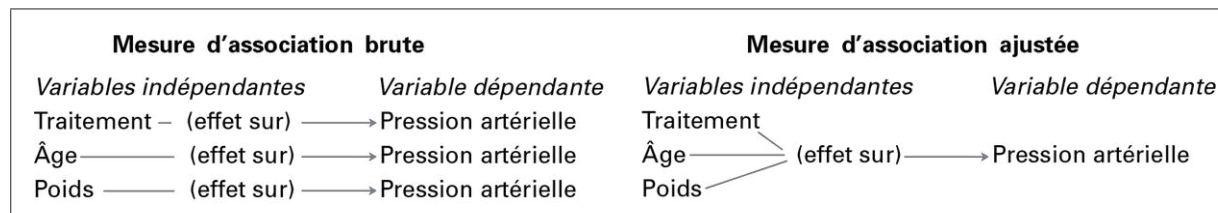
Tension artérielle diastolique (mm Hg) = 59,311 + 0,223 (poids en kilogrammes)

La valeur 59,311 correspond au symbole β_0 donné au modèle 12.1. Pratiquement, cette valeur n'est pas très utile dans l'interprétation des données, puisque personne ne pèse 0 kg. La valeur 0,223 correspond au symbole β_1 donné à ce même modèle. Cette valeur signifie que chaque kilogramme gagné entraîne une augmentation de la pression artérielle diastolique de 0,223 mm Hg. À la lumière de ce modèle de régression, une perte de 5 kg est associée à une diminution de la pression artérielle de 1,1 mm Hg.

12.2 L'utilité de l'analyse multivariée

Plusieurs phénomènes ou événements peuvent influencer une variable dépendante. On se retrouve alors avec plusieurs variables indépendantes. Par exemple, le traitement antihypertenseur n'est pas la seule variable indépendante à pouvoir influencer la pression artérielle. En effet, l'âge et le poids sont aussi des variables indépendantes susceptibles d'influer sur cette variable dépendante. L'analyse multivariée permet de mesurer la force des associations entre la variable dépendante et chacune des variables indépendantes. De plus, chacune des mesures d'association entre la variable dépendante et l'une des variables indépendantes tient compte de l'effet simultané des autres variables indépendantes considérées dans l'analyse. Dans certains contextes, la mesure d'association ainsi ajustée présente de grands avantages comparativement à la mesure d'association brute qui ne considère que l'effet d'une seule variable indépendante; la figure 12.2, à la page 246, illustre la distinction entre une **mesure d'association brute** et une **mesure d'association ajustée**.

Charles se rend à la pharmacie de la polyclinique Milo. Marie-Louise, la pharmacienne, lui remet son médicament antihypertenseur en lui rappelant l'importance d'être fidèle au traitement prescrit. Charles confie à Marie-Louise qu'il a de la difficulté à accepter de prendre des médicaments jusqu'à la fin de ses jours et qu'il préférerait avoir recours à un produit naturel. Il remet à la pharmacienne un dépliant publicitaire décrivant l'hypotense, un produit naturel qui ferait des miracles contre l'HTA. Au premier coup

FIGURE 12.2 ■ La distinction entre une mesure d'association brute et ajustée

d'œil, l'attention de la pharmacienne est attirée par l'affirmation suivante : « Dans une étude récente, la consommation d'hypotense a été associée à une diminution statistiquement significative de l'hypertension. Le risque d'être encore hypertendu six mois après le début de l'étude était moindre chez les patients hypertendus ayant pris de l'hypotense que chez les patients hypertendus ayant pris un placebo. »

Intéressée, Marie-Louise promet à Charles de s'informer sur ce produit et de lui faire part de son opinion le plus tôt possible. Le lendemain, elle repère rapidement sur Internet l'étude sur l'hypotense dont le dépliant publicitaire faisait mention. Il s'agit d'une étude expérimentale hasardisée au cours de laquelle 200 patients atteints d'HTA ont reçu l'hypotense, et 200 autres, un placebo; un programme d'aide a été offert aux participants des deux groupes afin qu'ils diminuent leur consommation de sel, fassent plus d'exercice et perdent du poids.

Les chercheurs ont d'abord calculé le **risque relatif brut** (RR_b) d'hypertension pour le groupe consommant l'hypotense par rapport au groupe recevant le placebo. Ce RR_b était de 0,8. L'intervalle de confiance à 95 % de ce rapport était de 0,7 à 0,9. Le RR_b étant inférieur à 1 et statistiquement significatif, puisque l'intervalle de confiance ne recouvrait pas la valeur 1, l'hypotense semblait plus efficace qu'un placebo contre l'HTA. Toutefois, l'analyse ne s'arrêtait pas là. Les chercheurs, craignant que les participants des deux groupes n'aient pas adhéré dans la même proportion aux recommandations du programme d'aide, ont calculé un **risque relatif ajusté** (RR_a). Ils ont inclus dans l'analyse les variables indépendantes suivantes : la prise de l'hypotense ou du placebo et les changements au cours de la période d'observation de la consommation de sel, du niveau de l'activité physique et du poids. Le RR_a a alors été de 0,9 avec un intervalle de confiance à 95 % de 0,8 à 1,2. L'intervalle recouvrant la valeur 1, il n'existait plus d'association statistiquement significative entre l'hypotense et l'HTA. Donc, lorsque les chercheurs ont pris en considération d'autres facteurs pouvant influencer la pression artérielle, ils ont observé que l'effet attribué à l'hypotense n'était pas réel. Les chercheurs ont soupçonné la présence d'un biais de confusion : les patients ayant reçu l'hypotense ont peut-être adhéré en plus grand nombre aux recommandations visant à modifier leur comportement que ceux ayant pris le placebo. La diminution de la tension artérielle chez ces individus serait alors attribuable moins à l'hypotense qu'au fait d'avoir diminué leur consommation de sel, augmenté leur activité physique ou perdu du poids. Marie-Louise informe donc Charles que l'hypotense n'est pas un produit miracle, puisqu'il ne diminue pas la pression artérielle plus que ne le fait un placebo.

12.2.1 Un rappel sur le biais de confusion

En tenant compte de l'effet de plusieurs variables indépendantes sur une variable dépendante, l'analyse multivariée permet de contrôler les biais de confusion. Comme cela est dit dans le chapitre 6, le biais de confusion se définit comme une erreur systématique induite par la présence d'un facteur associé à la fois à la variable dépendante et à la variable indépendante. Pour qu'il introduise un biais de confusion, ce facteur ne doit pas appartenir à la chaîne causale entre ces deux variables. La présence d'un tel facteur de confusion est à l'origine d'une erreur dans la mesure d'association brute. Lorsqu'il y a confusion dans les données, la mesure d'association brute s'avère biaisée et incorrecte.

Deux possibilités s'offrent alors aux chercheurs : présenter les mesures d'association spécifiques pour chacune des strates de la **variable confondante** ou présenter une mesure d'association ajustée. La première démarche est pertinente lorsqu'il n'y a qu'une variable confondante, ce qui demeure rare. Dans tous les autres cas, il est plus commode de présenter une mesure d'association ajustée. Celle-ci est exempte de l'erreur due aux facteurs de confusion considérés et résume bien l'information contenue dans l'ensemble des strates. L'analyse multivariée permet de calculer cette mesure d'association ajustée.

EXEMPLE 12.2

Une étude à visée étiologique cas-témoins a pour objectif de mesurer l'association entre le poids du bébé à la naissance et le tabagisme chez la mère. Dans cette étude, le gain de poids de la mère pendant la grossesse est considéré comme un facteur de confusion potentiel, car il se trouve associé au tabagisme et peut influencer le poids du bébé à la naissance. Pour les besoins de l'exemple, on considère que le poids de la mère n'appartient pas à la chaîne causale entre le tabagisme chez celle-ci et le poids du bébé à la naissance. Les mères sont partagées en deux strates : celles dont le gain de poids pendant la grossesse a été égal ou supérieur à 10 kg (strate A) et celles chez qui il a été inférieur à 10 kg (strate B). Les résultats de l'étude sont présentés dans le tableau 12.1.

TABLEAU 12.1 ■ Les résultats d'une étude à visée étiologique cas-témoins

Strate A : Femmes ayant pris 10 kg ou plus				Strate B : Femmes ayant pris moins de 10 kg				Ensemble des femmes, sans égard au gain de poids			
	Fument	Ne fument pas	Total		Fument	Ne fument pas	Total		Fument	Ne fument pas	Total
Cas	80	352	432	Cas	85	104	189	Cas	165	456	621
Témoins	28	380	408	Témoins	68	236	304	Témoins	96	616	712
Total	108	732	840	Total	153	340	493	Total	261	1 072	1 333

$RC_s = 3,08$ (intervalle de confiance à 95% : de 1,96 à 4,86)

$RC_b = 2,84$ (intervalle de confiance à 95% : de 1,91 à 4,20)

$RC_h = 2,32$ (intervalle de confiance à 95% : de 1,76 à 3,07)

EXEMPLE 12.2 (suite)

Le rapport de cotes brut (RC_b) de 2,32 indique que les mères qui fument risquent de mettre au monde un bébé de faible poids à la naissance 2,3 fois plus que les mères qui ne fument pas. Le rapport de cotes spécifique (RC_s) pour la strate des mères ayant pris 10 kg ou plus étant de 3,08, et celui observé pour la strate des mères ayant pris moins de 10 kg étant de 2,84, on constate que les mesures d'association spécifiques pour chacune des strates ne sont que légèrement différentes. De plus, les intervalles de confiance des rapports de cotes spécifiques se recoupant, on ne peut conclure qu'il y a une différence significative entre les deux rapports de cotes; la différence s'explique peut-être alors par les fluctuations d'échantillonnage. Cependant, on s'attendrait à ce que le RC_b (que l'on peut se représenter comme une sorte de mesure moyenne) résume l'information que l'on trouve dans les strates et qu'il se situe par conséquent de 3,08 à 2,84. Or, ici, le RC_b (2,32) se situe à l'extérieur de l'étendue formée par les rapports de cotes spécifiques, ce qui signifie qu'il est biaisé.

12.2.2 Le contrôle des biais de confusion

Lorsque plusieurs variables risquent de biaiser la mesure d'association entre une variable dépendante et une variable indépendante, il est essentiel de procéder à l'analyse multivariée. En effet, dans ce cas, il devient à peu près impossible de stratifier les données, car chaque stratification fragmente le groupe initial. On se trouve donc rapidement devant une réduction des effectifs qui rend l'analyse presque impossible. L'analyse multivariée permet d'obtenir une mesure d'association entre la variable dépendante et la variable indépendante ajustée simultanément pour toutes les variables confondantes. Toutefois, comme nous le verrons à la section 12.2.3, avant d'utiliser une telle mesure ajustée, il faut évaluer s'il y a une modification de la mesure d'association dans les strates.

EXEMPLE 12.3

L'étude sur l'association entre le poids des bébés à la naissance et le tabagisme chez les mères illustre cette première fonction de l'analyse multivariée. Le gain de poids de la mère pendant la grossesse est une variable qui a introduit un biais de confusion dans la mesure de l'association. D'autres variables, comme l'âge de la mère, son état de santé et son poids au début de la grossesse peuvent également biaiser la mesure de l'association entre le poids du bébé à la naissance et le tabagisme chez la mère. L'analyse multivariée permet de contrôler simultanément toutes ces variables confondantes et d'obtenir une mesure ajustée exempte de biais de confusion.

12.2.3 La modification de la mesure d'association

Parfois, les mesures présentées dans chacune des strates sont très différentes. Il y a alors **modification** de la mesure d'association, c'est-à-dire qu'une variable vient modifier l'association entre la variable dépendante et la variable indépendante considérée. S'il existe des tests statistiques qui permettent de déterminer la présence

ou l'absence de modification, la conclusion de ces tests étant en général rapportée dans les publications, on peut aussi l'estimer en observant les intervalles de confiance. Si les intervalles de confiance calculés pour les mesures obtenues dans chacune des strates respectives ne se recoupent pas, c'est qu'il y a modification. Cependant, si ces intervalles de confiance se recoupent, on ne peut conclure qu'il n'y a pas de modification, cette façon de faire ayant une puissance statistique plus faible que l'utilisation d'un test statistique.

Il est généralement reconnu que la variable *sexe* modifie l'association entre l'âge et l'incidence des maladies cardiovasculaires. En effet, chez les hommes, l'incidence de ces maladies augmente graduellement avec l'âge, alors que chez les femmes, elle ne commence à augmenter qu'après la cinquantaine. Lorsqu'il y a modification, comme dans ce cas, il n'est pas approprié de calculer une mesure ajustée qui résumerait les mesures spécifiques d'association de toutes les strates, car ce faisant, on perdrait des informations précieuses. Il est en effet intéressant de savoir que les femmes sont en quelque sorte «protégées» des maladies cardiovasculaires jusqu'à la ménopause.

Lorsqu'il y a modification, deux possibilités s'offrent aux chercheurs : présenter les mesures d'association spécifiques ou tenir compte de la modification dans l'analyse

EXEMPLE 12.4

Marie-Louise a trouvé un autre article décrivant une étude sur le traitement de l'HTA. Au total, 200 hommes et 200 femmes ont participé à cette étude expérimentale hasardisée dont l'objectif consistait à mesurer l'effet de l'activité physique sur le contrôle de l'HTA. Au début de l'étude, tous les sujets étaient hypertendus et sédentaires. Puis, 200 participants (100 hommes et 100 femmes) ont fait de l'activité physique 3 fois par semaine pendant 1 an, alors que l'on s'est abstenu d'inciter les 200 autres à augmenter leur niveau d'activité. L'association entre l'activité physique et le contrôle de l'HTA a été mesurée en établissant les risques relatifs. Les chercheurs soupçonnaient une influence de la variable *sexe* sur l'association entre l'activité physique et l'HTA. Ils ont donc effectué une analyse stratifiée selon le sexe; les résultats obtenus sont présentés dans le tableau 12.2.

TABLEAU 12.2 ■ Les résultats d'une étude expérimentale hasardisée (données stratifiées selon le sexe)

Strate A : Hommes				Strate B : Femmes			
Activité physique	HTA contrôlée	HTA non contrôlée	Total	Activité physique	HTA contrôlée	HTA non contrôlée	Total
Oui	80	20	100	Oui	58	42	100
Non	30	70	100	Non	41	59	100

$RR_s = 80/100 \div 30/100 = 2,67$ (intervalle de confiance à 95% : 1,95 à 3,65)

$RR_s = 58/100 \div 41/100 = 1,41$ (intervalle de confiance à 95% : 1,06 à 1,89)

EXEMPLE 12.4 (suite)

Ils ont obtenu un risque relatif spécifique (RR_s) de 2,67 pour les hommes (strate A) et un autre de 1,41 pour les femmes (strate B). De plus, les intervalles de confiance de ces RR_s ne se recoupaient pas. L'association entre l'activité physique et la diminution de l'HTA serait donc plus forte chez les hommes que chez les femmes. Afin de mettre cette différence bien en évidence, les chercheurs ont décidé de ne pas résumer l'association entre l'activité physique et le contrôle de l'HTA à l'aide d'une mesure d'association ajustée selon le sexe. La mesure ajustée aurait en effet nivelé la différence, et la réalité s'en serait trouvée inadéquatement reflétée.

multivariée. Souvent, si la modification est considérable et qu'elle permet de mieux décrire les résultats de l'étude, les chercheurs préféreront présenter les mesures spécifiques, comme dans le tableau 12.2. Prendre en considération la modification de la mesure d'association dans l'analyse multivariée est complexe et nécessite l'ajout d'un élément appelé « terme d'interaction ». Il n'en sera pas question ici ; le lecteur pourra se reporter aux ouvrages spécialisés² pour approfondir le sujet.

12.3 Les méthodes d'analyse multivariée

Les fonctions de l'**analyse multivariée** consistent à contrôler les biais de confusion en ce qui concerne les variables incluses dans le modèle, et à mesurer l'effet simultané de plusieurs variables indépendantes sur une variable dépendante. Cependant, aussi pointue que puisse être l'analyse multivariée, elle ne peut corriger les faiblesses d'une étude qui a été menée de façon inadéquate. L'appréciation d'une étude doit d'abord porter sur les méthodes utilisées par les chercheurs pour vérifier leur hypothèse de recherche. En effet, si l'analyse multivariée permet de contrôler un biais de confusion, elle ne peut corriger un biais d'information ou un biais de sélection. De plus, elle ne peut pallier un manque de puissance statistique, une taille d'échantillon trop petite ou l'absence de mesure d'une variable importante.

Il existe plusieurs méthodes d'analyse multivariée. Chacune d'entre elles repose sur un **modèle mathématique** et sur un ensemble de conditions d'application. Un modèle mathématique est constitué d'expressions mathématiques qui représentent la variable dépendante en fonction de plusieurs variables indépendantes. Les modèles mathématiques sont donc des représentations plus ou moins simplifiées de l'association qui existe entre la variable dépendante et les variables indépendantes. Utilisés à différentes fins, ces modèles peuvent servir à décrire un phénomène, à l'expliquer ou à le prédire. Un modèle multivarié peut aussi être employé pour remplacer un ensemble de variables indépendantes par une seule variable. Cette nouvelle variable se nomme **score de propension**. Ce score s'avère particulièrement utile lorsque le phénomène est un événement rare.

Les méthodes d'analyse multivariée les plus fréquemment utilisées en épidémiologie sont la régression linéaire multiple, la régression logistique et la régression à effet proportionnel ou régression de Cox. Le choix d'une de ces méthodes est fonction du type de variable dépendante et des mesures d'intérêt, comme le résume le tableau 12.3.

TABLEAU 12.3 ■ Les méthodes d'analyse multivariée les plus utilisées

Variable dépendante	Mesure d'intérêt	Méthode d'analyse multivariée
Quantitative continue	Coefficient de régression (β)	Régression linéaire multiple
Catégorielle	Rapport de cotes (RC)	Régression logistique
Catégorielle	Rapport d'incidence (RI)	Régression à effet proportionnel (de Cox)

À retenir

L'analyse multivariée permet de contrôler les biais de confusion et de connaître l'effet simultané de plusieurs variables indépendantes sur une variable dépendante.

12.3.1 La régression linéaire multiple

Comme on l'a vu à la section 12.1.1, la relation entre une variable indépendante et une variable dépendante quantitative continue peut être mesurée à l'aide d'une régression linéaire simple. Ainsi en est-il pour décrire la relation entre la pression artérielle diastolique et le poids. Toutefois, on pourrait aussi considérer plusieurs variables indépendantes ($x_1, x_2, x_3, \dots, x_i$) en effectuant une **régression linéaire multiple**. Cette dernière s'exprime par le modèle mathématique suivant.

Régression linéaire multiple (modèle mathématique 12.2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_i x_i$$

Les éléments de droite du modèle 12.2 correspondent aux variables indépendantes: ainsi, chaque symbole β_i représente le coefficient de régression associé à une variable indépendante x_i . Théoriquement, un modèle de régression linéaire multiple peut considérer n'importe quel nombre de variables indépendantes. Ce type de modèle peut donc être utilisé pour expliquer les variations de la pression artérielle diastolique en tenant compte de plusieurs variables indépendantes.

EXEMPLE 12.5

Dans une étude menée au Québec, les chercheurs ont mis au point le modèle mathématique plus complexe suivant.

Pression artérielle diastolique (mm Hg) = 58,214 + 0,20 (poids en kilogrammes) - 0,18 (heures d'activité physique par semaine)

Dans ce modèle, l'effet de l'activité physique sur la pression artérielle diastolique est pris en compte. Selon ce modèle, chaque heure d'activité physique hebdomadaire correspond à une diminution de la pression artérielle diastolique de 0,18 mm Hg.

Les résultats de la régression linéaire multiple présentés dans le tableau 12.4 portent sur les facteurs de risque de l'hypercholestérolémie. Dans ce tableau, les coefficients de régression (β) sont présentés pour chacune des variables indépendantes et sont accompagnés d'une valeur p.

TABEAU 12.4 ■ Les résultats de la régression linéaire multiple

Variable	β (coefficient de régression)	Valeur p
Âge (années)	0,022	< 0,001
Sexe*	-0,198	0,17
IMC** (kg/m ²)	0,046	< 0,004
Apport alimentaire en cholestérol (mg/jour)	< 0,001	0,32

Coefficient de détermination multiple (R^2) total = 0,20

* Le sexe masculin a servi de référence.

** L'IMC (indice de masse corporelle) se calcule en divisant le poids (kg) par la taille au carré (m²).

Chacune des informations que contient ce tableau peut nous aider à mieux comprendre les variables qui influencent la cholestérolémie. Chacun des coefficients de régression nous renseigne sur l'importance de la relation entre une variable indépendante et la variable dépendante. La valeur p indique la probabilité d'observer un coefficient de cette ampleur (ou encore plus grand) si l'hypothèse nulle (H_0) est vraie. Ainsi, le β de 0,022 montre qu'il y a une relation entre l'âge et la cholestérolémie et que cette relation est statistiquement significative ($p < 0,001$). Le β de 0,046 obtenu pour l'IMC ainsi que la valeur p qui l'accompagne ($p < 0,004$) indiquent que l'IMC est associé à la cholestérolémie. Par contre, ni le sexe ni l'apport alimentaire en cholestérol ne sont significativement liés à la cholestérolémie.

De plus, une valeur de R^2 accompagne le tableau. Le R^2 , aussi appelé **coefficient de détermination multiple**, est le carré du coefficient de corrélation. Le R^2 indique quelle proportion de la variation de la variable dépendante est attribuable à l'ensemble des variables indépendantes incluses dans le modèle. Si R^2 égalait 1, le modèle prédirait sans erreur la variable dépendante pour chaque individu. Les variables incluses dans le modèle expliqueraient 100 % de la variation; c'est évidemment là une situation que l'on ne trouve jamais dans la littérature scientifique. Ainsi, un modèle dont le R^2 est de 0,84 fournira plus de renseignements sur la variation de la variable dépendante qu'un modèle où le R^2 a une valeur de 0,20.

Dans le tableau 12.4, le R^2 s'élève à 0,20. Cela signifie que les variables indépendantes incluses dans le modèle de régression sont à l'origine de 20 % de la variation de la variable dépendante. Les variables indépendantes *âge*, *sexe*, *indice de masse corporelle (IMC)* et *apport alimentaire en cholestérol* expliquent donc 20 % de la variation de la cholestérolémie observée chez les individus. D'autres variables fourniraient une explication de la proportion de la variation qui échappe au présent modèle, soit 80 %; ces variables pourraient être l'hérédité, l'activité physique, le tabagisme, etc.

Pour le clinicien, le R^2 sert surtout à déterminer dans quelle mesure un modèle peut être utilisé pour prédire le risque que présentent pour la santé d'un individu les variables analysées. Par exemple, les variables indépendantes considérées dans le modèle du tableau 12.4, soit l'âge, le sexe, l'IMC et l'apport alimentaire en cholestérol, expliquent une part relativement faible de la variation de la cholestérolémie.

À retenir

- La régression linéaire multiple permet d'analyser l'association entre une variable dépendante et plusieurs variables indépendantes lorsque la variable dépendante est une variable quantitative continue.
- Les coefficients β_i ($i > 0$) de la régression linéaire nous renseignent sur l'importance de la relation entre une variable indépendante i et la variable dépendante.
- Lorsque plusieurs variables indépendantes sont considérées, les coefficients β_i ($i > 0$) obtenus sont ajustés pour les variables indépendantes incluses dans le modèle.
- Le R^2 indique la proportion de la variation de la variable dépendante qui est attribuable à l'ensemble des variables indépendantes incluses dans le modèle.

12.3.2 La régression logistique

La **régression logistique** constitue une technique d'analyse multivariée utilisée lorsque la mesure d'association est un rapport de cotes et que la variable dépendante est catégorielle et dichotomique, c'est-à-dire qu'elle ne comporte que deux valeurs possibles.

Comme on l'a vu au début du présent chapitre, Charles a pris un médicament offert en vente libre pour soulager les symptômes du rhume. Lisa se rappelle avoir lu une étude dans laquelle des chercheurs se sont penchés sur l'hypothèse voulant que l'automédication puisse diminuer le nombre de visites chez le médecin. Les chercheurs ont d'abord étudié l'association entre le fait d'avoir consommé un médicament sans l'avis d'un médecin dans les deux jours précédant l'enquête et celui d'avoir consulté un médecin dans les deux semaines précédentes. Ils en sont arrivés à une association statistiquement significative entre la consommation d'un médicament offert en vente libre et la consultation d'un médecin. En effet, le rapport de cotes brut était de 1,51, et les limites de l'intervalle de confiance à 95 % de 1,36 à 1,68. Cette mesure brute donnait à penser que les personnes qui avaient consulté récemment un médecin prenaient en plus grande proportion que les autres des médicaments offerts en vente libre. Ces résultats un peu étonnants ont motivé les chercheurs à procéder à une analyse multivariée.

À l'aide d'une régression logistique, ils ont réexaminé l'association entre la consommation de médicaments offerts en vente libre et le recours aux soins médicaux, en tenant compte, ici, d'autres variables reconnues pour être associées à la fois à l'utilisation de soins médicaux et à la consommation de ce type de médicaments : l'âge, le sexe et le nombre de problèmes de santé.

La régression logistique a donné le modèle mathématique suivant.

Consommation de médicaments offerts en vente libre = $\beta_0 + \beta_1$ (consultation d'un médecin) + β_2 (âge) + β_3 (sexe) + β_4 (nombre de problèmes de santé)

Au moyen d'une transformation algébrique, les β obtenus par régression logistique peuvent être facilement transformés en rapports de cotes, soit des rapports de cotes ajustés (RC_a).

Les résultats de l'analyse effectuée par les chercheurs sont présentés dans le tableau 12.5.

TABLEAU 12.5 ■ Les résultats de l'analyse

Variable	Rapport de cotes	Intervalle de confiance à 95 %
Consultation d'un médecin		
Non	1,00	-
Oui	0,80	0,65 à 0,95
Âge (années)		
15-24	1,00	-
25-34	1,19	1,03 à 1,37
35-44	1,36	1,17 à 1,59
45-54	1,35	1,15 à 1,60
55-64	1,11	0,93 à 1,33
≥ 65	0,94	0,78 à 1,14
Sexe		
Hommes	1,00	-
Femmes	1,49	1,37 à 1,61
Nombre de problèmes de santé		
Aucun	1,00	-
1	3,30	2,96 à 3,67
> 1	6,69	5,99 à 7,48

La mesure d'association ajustée, soit un rapport de cotes de 0,80 avec un intervalle de confiance à 95 % de 0,65 à 0,95, est très différente de la mesure d'association brute (1,51). En fait, à la lumière de ces données, on constate que les personnes qui ont récemment consulté un médecin prennent en moins grande proportion que les autres des médicaments offerts en vente libre.

Comment expliquer que deux analyses portant sur la même population conduisent à des résultats contradictoires? C'est que, dans l'analyse brute, une seule variable indépendante – la consultation d'un médecin – a été considérée, alors que l'analyse multivariée a pris en compte quatre variables. Ainsi, on peut envisager que le nombre de problèmes de santé que présente un individu soit associé à la fois à la consommation de médicaments offerts en vente libre et à la consultation d'un médecin. Cet exemple illustre la façon dont l'analyse multivariée permet de contrôler un tel biais de confusion.

À retenir

- La régression logistique permet d'analyser l'association entre une variable dépendante et une ou plusieurs variables indépendantes, lorsque la variable dépendante est une variable dichotomique.
- Les coefficients β_i ($i > 0$) de la régression logistique se transforment pour obtenir des rapports de cotes.
- Lorsque plusieurs variables indépendantes sont considérées, les rapports de cotes obtenus sont ajustés pour les variables indépendantes incluses dans le modèle.

12.3.3 La régression à effet proportionnel ou régression de Cox

Lorsqu'on cherche à estimer le risque que survienne une maladie ou un événement en tenant compte du temps écoulé entre l'exposition à la variable indépendante et la survenue de la maladie ou de l'événement — c'est-à-dire lorsqu'on veut calculer des rapports de taux d'incidence —, on peut utiliser la **régression à effet proportionnel** ou **régression de Cox**, qui sert notamment à l'analyse des études de cohorte(s). Ainsi, si l'événement est la mort du sujet, le temps écoulé depuis le début de l'étude jusqu'à la mort du sujet représente sa survie.

L'objectif thérapeutique que vise Lisa est la baisse de la pression artérielle diastolique de son patient, car une pression artérielle élevée, qu'elle soit diastolique ou systolique, augmente le risque de problèmes cardiovasculaires. Une étude démontrant ce risque associé à la pression artérielle systolique a été menée à Québec, il y a plusieurs années.

EXEMPLE 12.6

En 1974, des chercheurs ont recruté 4 828 hommes alors âgés de 35 ans à 64 ans³. La pression artérielle systolique, le tabagisme, la cholestérolémie et le niveau de scolarité de chaque participant ont été mesurés au début de l'étude. Des informations sur la survie de ces hommes ont été recueillies pendant 12 ans. De janvier 1974 à janvier 1986, 417 participants sont décédés. Les chercheurs ont calculé des rapports de taux de mortalité.

EXEMPLE 12.6 (suite)

Pour mesurer l'effet de la pression artérielle systolique sur la mortalité, les chercheurs auraient-ils pu simplement se contenter de comparer deux taux d'incidence, c'est-à-dire la mortalité chez les participants ayant une pression artérielle systolique élevée et celle chez les participants dont la pression artérielle systolique est normale? Non, puisqu'ils auraient alors obtenu une mesure d'association brute potentiellement biaisée, de nombreuses variables pouvant fausser l'association entre la pression artérielle systolique et la mortalité. L'âge, par exemple, est une variable confondante, puisque tant le risque d'être atteint d'HTA que celui de décéder augmentent avec l'âge.

Afin d'obtenir une mesure non biaisée de l'association entre la pression artérielle systolique et la mortalité, les chercheurs ont dû calculer un rapport de taux d'incidence ajusté (RTI_a) pour les variables confondantes. Ils ont pour ce faire utilisé une régression à effet proportionnel ou régression de Cox. Il s'agit d'un modèle de régression qui permet de tenir compte du calcul des personnes-temps. Grâce à une transformation algébrique, les coefficients de régression (β) obtenus à l'aide d'une régression de Cox peuvent être transformés en rapports de taux d'incidence ou de mortalité.

Le tableau 12.6 présente quelques résultats obtenus par l'équipe de chercheurs québécois.

TABLEAU 12.6 ■ Les résultats d'une étude*

Pression artérielle systolique (mm Hg)	Nombre de personnes-année	Nombre d'hommes	Nombre de décès	Rapport de taux de mortalité toutes causes confondues	Intervalle de confiance à 95 %
< 125	9 909	931	49	1,0	-
125-132	10 632	985	59	1,1	0,7 à 1,5
133-140	9 163	863	69	1,4	0,9 à 1,9
141-152	9 827	928	93	1,5	1,1 à 2,2
> 152	8 904	868	147	2,0	1,5 à 2,8

* Cela inclut les rapports de taux de mortalité, toutes causes confondues, ajustés selon l'âge, la cholestérolémie, le tabagisme et le niveau de scolarité en fonction du quintile de pression artérielle systolique.

Source: Adapté de Dagenais, G.R., Z. Ahmed, N.M. Robitaille et autres. «Total and coronary heart disease mortality in relation to major risk factors – Quebec cardiovascular study», *Canadian Journal of Cardiology*, 1990, vol. 6, n° 7, p. 59-65. Reproduit avec l'autorisation de l'éditeur.

On note que les rapports de taux de mortalité, toutes causes confondues, sont plus élevés dans les quintiles supérieurs de pression artérielle. L'association entre la pression artérielle systolique et la mortalité est statistiquement significative pour les deux quintiles supérieurs; cette association n'est pas biaisée par les variables confondantes incluses dans le modèle. Les valeurs brutes (biaisées) pour chacune des catégories de pression artérielle 125-132, 133-140, 141-152, > 152 sont respectivement 1,1, 1,5, 1,9 et 3,3. Les mesures brutes auraient donc surestimé le risque de décès.

À retenir

- La régression à effet proportionnel ou régression de Cox permet d'estimer le risque que survienne une maladie ou un événement en tenant compte du temps écoulé entre l'exposition à la variable indépendante et la survenue de la maladie ou de l'événement, tout en contrôlant les variables confondantes.
- Les coefficients β_i ($i > 0$) se transforment pour estimer des rapports de taux d'incidence ou de mortalité.
- Ce type de régression permet d'obtenir une estimation ajustée du rapport de taux d'incidence ou de mortalité.

En résumé

Après avoir défini les variables dépendantes et indépendantes, on a vu dans ce chapitre que l'analyse multivariée permet de contrôler les biais de confusion et de connaître l'effet simultané de plusieurs variables indépendantes sur une variable dépendante. On a établi la distinction entre les notions de confusion et de

modification de la mesure d'association. Enfin, on a souligné que la méthode d'analyse multivariée à utiliser (régression linéaire multiple, régression logistique, régression à effet proportionnel ou régression de Cox) est fonction des caractéristiques de la variable dépendante et de la mesure d'association d'intérêt.

Exercices



Voir la section complète des exercices sur le Web.

Indiquez si les énoncés suivants sont vrais ou faux.

1. Les méthodes d'analyse multivariée permettent de prendre en compte l'effet simultané de plusieurs variables indépendantes sur une variable dépendante.
2. En général, lorsque la mesure brute est plus élevée ou moins élevée que la mesure spécifique la plus élevée ou la moins élevée représentée dans les strates, c'est qu'un biais de confusion influe sur la mesure brute.
3. Lorsque la variable dépendante est catégorielle et dichotomique et que la mesure d'intérêt est un rapport de cotes, la régression logistique s'avère la plus appropriée parmi les méthodes d'analyse multivariée.
4. La régression linéaire multiple permet d'ajuster la mesure d'association entre chacune des variables indépendantes retenues et la variable dépendante, lorsque cette dernière est quantitative et continue.
5. Le choix de la méthode d'analyse multivariée dépend des variables indépendantes.
6. La régression logistique permet d'estimer un rapport de cotes.
7. La régression à effet proportionnel ou régression de Cox s'applique aux études cas-témoins.

Pour les réponses, voir le corrigé à la fin du manuel.

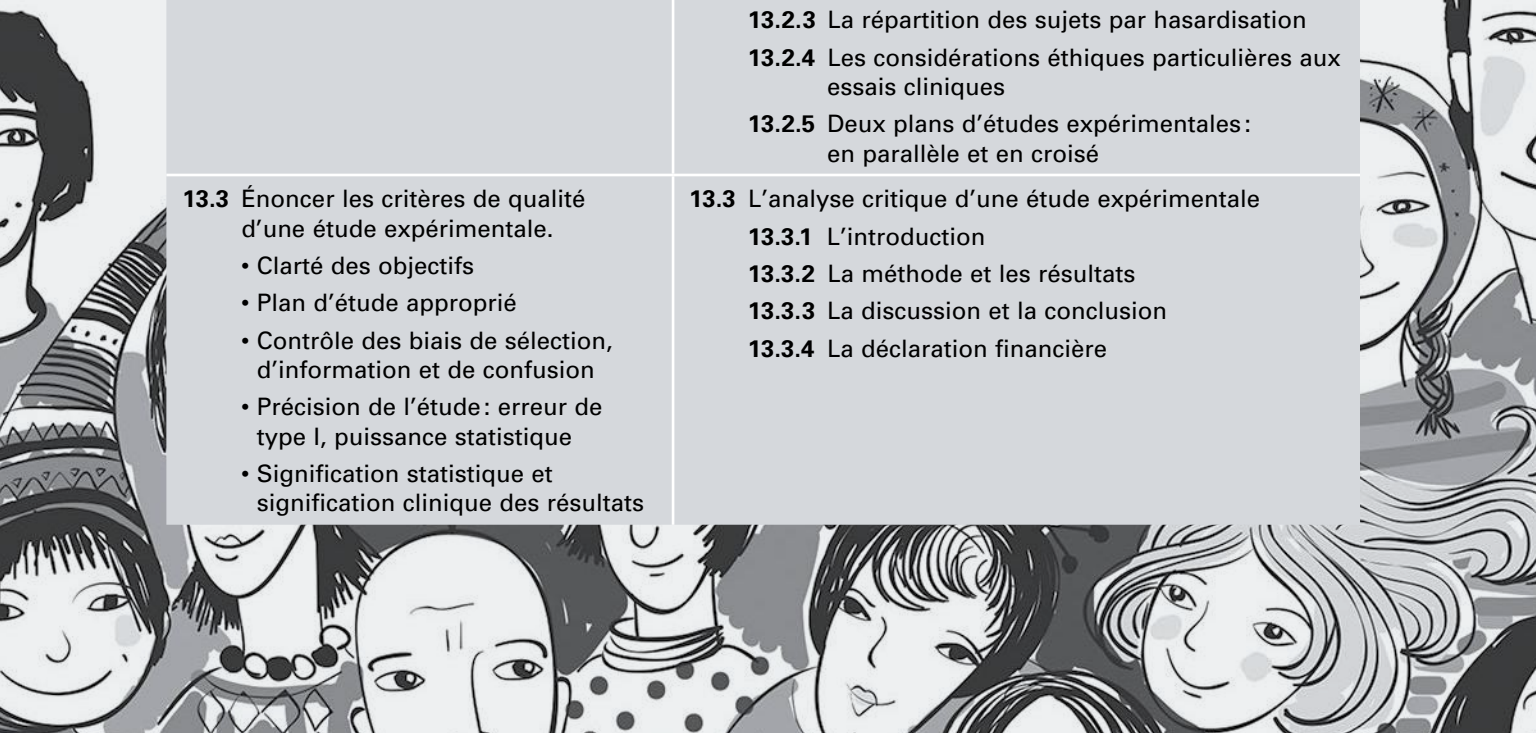
Collaboration à la 3^e édition :

Danielle Laurin

Au terme de ce chapitre, vous pourrez :

- réaliser la lecture critique d'articles scientifiques portant sur les études expérimentales selon une méthode structurée et systématique ;
- juger de la pertinence de modifier votre pratique à la lumière des résultats d'une publication scientifique portant sur les études expérimentales.

Objectifs spécifiques	Contenu
13.1 Décrire les facteurs qui peuvent influencer la réponse à un traitement ou à une intervention.	13.1 Les facteurs susceptibles d'expliquer la réponse à un traitement ou à une intervention <ul style="list-style-type: none"> 13.1.1 L'histoire naturelle de la maladie 13.1.2 L'effet placebo 13.1.3 L'effet de l'attention accordée au sujet 13.1.4 L'effet réel d'un traitement ou d'une intervention
13.2 Définir une étude expérimentale et ses variantes.	13.2 Qu'est-ce qu'une étude expérimentale? <ul style="list-style-type: none"> 13.2.1 La définition et les étapes de la démarche 13.2.2 La sélection des sujets 13.2.3 La répartition des sujets par hasardisation 13.2.4 Les considérations éthiques particulières aux essais cliniques 13.2.5 Deux plans d'études expérimentales : en parallèle et en croisé
13.3 Énoncer les critères de qualité d'une étude expérimentale. <ul style="list-style-type: none"> • Clarté des objectifs • Plan d'étude approprié • Contrôle des biais de sélection, d'information et de confusion • Précision de l'étude : erreur de type I, puissance statistique • Signification statistique et signification clinique des résultats 	13.3 L'analyse critique d'une étude expérimentale <ul style="list-style-type: none"> 13.3.1 L'introduction 13.3.2 La méthode et les résultats 13.3.3 La discussion et la conclusion 13.3.4 La déclaration financière





Mise en situation

À la polyclinique Milo, l'omnipraticienne Lisa voit entrer sa grand-mère. La vieille dame, qui vient la consulter pour une verrue logée sous le pied, refuse de recevoir quelque traitement que ce soit sans l'accord de sa petite-fille. Certaines personnes lui ont suggéré d'utiliser l'un de ces médicaments offerts en vente libre dans les pharmacies tandis que d'autres lui ont recommandé un médicament homéopathique.

« À ton avis, Lisa, quel est le meilleur traitement ? » lui demande sa grand-mère.

Pertinence clinique

Devant l'abondance des publications scientifiques, les professionnels de la santé sont de plus en plus amenés à appuyer leur conduite sur des preuves scientifiques; ils doivent donc savoir critiquer les études expérimentales, car ce sont elles qui offrent les preuves les plus solides quant à l'efficacité d'un traitement, d'une intervention ou d'un programme. Elles constituent de ce fait l'une des principales sources d'information.

13.1 Les facteurs susceptibles d'expliquer la réponse à un traitement ou à une intervention

Quatre facteurs sont susceptibles d'expliquer la réponse à un traitement ou à une intervention: l'histoire naturelle de la maladie, l'effet placebo, l'effet de l'attention accordée au sujet et l'effet réel du traitement ou de l'intervention.

13.1.1 L'histoire naturelle de la maladie

L'évolution d'une maladie suit habituellement un parcours précis, parfois inconnu ou tellement complexe qu'il est difficile de le prédire. En l'absence d'intervention, la maladie peut se résorber graduellement jusqu'à disparaître; elle peut aussi se stabiliser ou progresser. L'évolution de la maladie est indépendante du traitement, mais elle peut être confondue avec l'effet de celui-ci. C'est pourquoi l'évaluation d'un traitement ne peut se faire correctement que si l'on constitue deux groupes: l'un recevant le traitement et l'autre servant de témoin. On pourra ainsi distinguer l'évolution naturelle de la maladie de l'effet du traitement.

EXEMPLE 13.1

La verrue plantaire de la grand-mère de Lisa peut guérir sans être traitée. En effet, selon certaines études, environ 10 % des verrues plantaires disparaissent un mois après être apparues. Dans de tels cas, si l'on menait une étude sur le traitement de la verrue plantaire, sa disparition pourrait s'expliquer tout aussi bien par son histoire naturelle que par l'effet du traitement. On ne peut donc tirer de conclusion sur l'efficacité réelle du traitement.

13.1.2 L'effet placebo

Le **placebo** est par définition un traitement factice. Ce terme s'applique généralement au médicament utilisé, à l'effet observé ou au groupe témoin. Lorsqu'il désigne un médicament, il signifie que la substance administrée n'a aucun effet pharmacologique reconnu. Il peut tout de même soulager, voire guérir, certaines affections. C'est ce qu'on appelle l'**effet placebo**. Dans de tels cas, c'est la croyance en l'efficacité du traitement proposé qui déclenche un processus de soulagement ou de guérison¹.

L'article clé de Beecher² portant sur des études effectuées sans groupe témoin, mais en aveugle rapporte que, dans l'une de ces études, 31 % des patients affirment avoir été soulagés grâce à un traitement administré après une opération alors qu'il s'agissait en fait d'un placebo. Dans le même article, l'auteur observe que 58 % des personnes victimes du mal de mer disent que leur état s'est amélioré à la suite de l'administration d'un placebo.

13.1.3 L'effet de l'attention accordée au sujet

Dans le cadre d'un programme visant à améliorer la productivité des employés, la compagnie Western Electric de Chicago a procédé, il y a près d'un demi-siècle, à une évaluation de l'impact de l'éclairage des sites de travail. Les employés ont participé activement aux consultations, et l'éclairage a été amélioré en tenant compte de leurs suggestions. Les résultats se sont révélés impressionnants : la productivité s'est accrue d'un très fort pourcentage. Afin de savoir si le nouvel éclairage était véritablement à l'origine de l'augmentation de la productivité, on y a apporté plusieurs changements, allant jusqu'à remettre en fonction les anciens systèmes. La productivité est restée la même. L'attention accordée aux employés expliquait donc l'amélioration de leur productivité.

Dans toutes les disciplines des sciences de la santé, on reconnaît que l'attention donnée au sujet entraîne des effets ; c'est ce qu'on appelle l'**effet de l'attention accordée au malade**. Le médecin tout comme l'infirmier ou l'ergothérapeute accordent beaucoup d'attention à la personne souffrante à laquelle ils administrent des soins. Les effets apparemment thérapeutiques de cette attention doivent donc être estimés et distingués lorsqu'on veut évaluer les effets d'un traitement.

EXEMPLE 13.2

Les sujets d'une étude portant sur l'association entre l'alimentation et les maladies cardiovasculaires sont interrogés chaque année sur leur régime alimentaire. Cette attention peut les amener à modifier graduellement leurs habitudes. Pour équilibrer les effets de l'attention accordée aux sujets et ceux du traitement évalué, on accordera la même attention aux deux groupes, soit celui recevant le placebo et celui recevant le traitement. Il sera alors possible de distinguer l'influence du traitement de l'attention accordée aux sujets sur leur état de santé.

13.1.4 L'effet réel d'un traitement ou d'une intervention

Afin d'être en mesure d'évaluer les répercussions réelles d'un traitement ou d'une intervention, les chercheurs doivent tout mettre en œuvre pour maîtriser ou exclure l'effet de l'histoire naturelle de la maladie, l'effet placebo et l'effet de l'attention accordée aux sujets. Ces deux derniers effets sont souvent difficiles à départager.

Dans une étude expérimentale, on utilise un groupe témoin pour estimer l'effet de l'intervention ou du traitement étudié en contrôlant les autres facteurs. Tant dans le groupe qui reçoit le traitement que dans le groupe témoin, les effets de l'histoire naturelle et de l'attention accordée au sujet se manifestent. La seule caractéristique qui distingue le second groupe du premier, c'est qu'il reçoit un placebo. Les différences entre les deux groupes ne peuvent donc être imputables qu'à l'intervention ou au traitement étudié.

EXEMPLE 13.3

L'étude HYVET³ (*Hypertension in the Very Elderly Trial*) est une étude expérimentale de grande envergure visant à évaluer les effets de la baisse de pression artérielle chez les personnes hypertendues âgées de 80 ans et plus. En 2001, 3 845 sujets ont été répartis au hasard et l'étude a été réalisée en double aveugle. Les sujets ont reçu soit un placebo, soit un traitement à base d'indapamide, éventuellement associé au péridopril, pour atteindre l'objectif de mesure de pression artérielle de 150/80 mm Hg. On a exercé un contrôle rigoureux des variables autres que celles liées au traitement. On a pu ainsi établir avec certitude l'efficacité du traitement indapamide et péridopril. L'administration de ce traitement a conduit à une diminution de la mortalité totale, de la mortalité par accident vasculaire cérébral et des insuffisances cardiaques.

À retenir

Afin de déterminer l'effet réel d'un traitement ou d'une intervention, il importe de contrôler les effets de l'histoire naturelle de la maladie, du placebo et de l'attention accordée au sujet.

13.2 Qu'est-ce qu'une étude expérimentale ?

13.2.1 La définition et les étapes de la démarche

L'**étude expérimentale** est considérée comme le modèle d'étude le plus susceptible de prévenir les biais et de produire les meilleures preuves sur le plan scientifique. De plus, il est le plus fiable pour établir un lien de causalité entre une exposition et un événement d'intérêt. Dans ce modèle d'étude, les sujets sont assignés ou non au traitement expérimental. Lorsque la distribution des sujets dans les groupes est déterminée aléatoirement (au hasard), on parle d'étude expérimentale «hasardisée». Lorsqu'il est impossible de recourir à cette méthode, il s'agit plutôt d'une étude expérimentale «non hasardisée».

Il arrive aussi que les sujets ne puissent être individuellement assignés au traitement étudié par risque de contamination entre les groupes. Dans ce cas, on doit assigner les groupes — et non les sujets — au traitement étudié. Il s'agit alors d'une **étude quasi expérimentale** parce que l'assignation du traitement n'est pas faite selon chaque individu, mais selon un groupe d'individus. Tout comme l'étude expérimentale, l'étude quasi expérimentale peut être hasardisée si l'assignation des groupes au traitement se fait de façon aléatoire.

EXEMPLE 13.4

Au cours de l'accouchement, il arrive que le bébé à naître présente des signes inquiétants qui contraignent l'équipe de soins à intervenir rapidement. Leveno et ses collègues⁴ ont comparé l'efficacité du monitoring fœtal universel (offert à toutes les femmes) avec celle du monitoring sélectif (offert aux femmes à risque seulement) pour déceler les signes de détresse chez le fœtus. Or, bien que la répartition par hasardisation des femmes dans les groupes ait été possible, elle ne permettait pas d'assurer le respect des consignes. En effet, les médecins qui croyaient en l'efficacité du monitoring universel auraient eu tendance à l'utiliser auprès de leurs patientes à la moindre difficulté, qu'elles soient ou non à risque, et cela, en dépit du fait que leur décision de participer à l'étude aurait normalement dû les inciter à ne pas tenir compte de leurs convictions. Afin de contourner ce problème, les chercheurs ont plutôt choisi de limiter le nombre d'appareils de monitoring. Pendant un mois donné, le petit nombre d'appareils forçait les médecins à ne les utiliser qu'auprès des femmes à risque, alors que, le mois suivant, leur nombre était suffisant pour offrir le monitoring universel à toutes les femmes. Cette étude quasi expérimentale a permis de conclure que le monitoring universel s'avérait inutile.

Bien que l'expression «étude expérimentale» renvoie à un plan d'étude précis, les études de ce type possèdent différentes appellations selon le contexte. En effet, une étude expérimentale qui a pour but d'évaluer l'efficacité d'un nouveau médicament ou d'une nouvelle intervention est communément appelée «essai clinique». Une étude portant sur l'efficacité d'une intervention dans une communauté, une campagne de vaccination par exemple, se nomme «essai communautaire».

L'étude expérimentale respecte une démarche précise. Le chercheur formule d'abord une hypothèse ; celle-ci pourrait énoncer, par exemple, que le traitement homéopathique

de la verrue plantaire est plus efficace que le placebo. Il détermine ensuite la population cible à laquelle il souhaite généraliser ses résultats. Puis, il définit la population à l'étude à partir de laquelle il pourra constituer un échantillon. Il établit alors des critères d'inclusion ou d'exclusion, posant par exemple que les personnes de 6 ans à 59 ans — ayant au moins une verrue plantaire — seront incluses et que les femmes enceintes et celles qui allaitent seront exclues. Ensuite, il constitue l'échantillon et il forme les groupes par hasardisation, puis il détermine les critères de jugement qui serviront à mesurer les effets du traitement. Finalement, il choisit des méthodes de prévention et de contrôle des biais telles que la technique en aveugle, l'évaluation des interventions autres que le traitement évalué (co-interventions) ou l'adoption de stratégies pour éviter de perdre de vue des sujets pendant l'étude. Il soumet son projet à un comité d'éthique et de recherche pour ensuite, après approbation, effectuer l'étude, en analyser les résultats et rédiger un manuscrit en vue d'une publication ou d'un rapport. Voici les 12 étapes de la démarche utilisée pour réaliser une étude expérimentale.

1. Formuler l'hypothèse de recherche.
2. Définir la population cible.
3. Définir la population à l'étude.
4. Établir les critères d'inclusion et d'exclusion.
5. Constituer un échantillon.
6. Former les groupes par hasardisation.
7. Établir les critères de jugement qui permettront de mesurer les résultats de l'étude.
8. Déterminer les méthodes de prévention et de contrôle des biais.
9. Faire approuver l'étude par un comité d'éthique et de recherche.
10. Effectuer la collecte des données de l'étude.
11. Analyser les résultats de l'étude.
12. Rédiger un manuscrit pour une publication ou un rapport.

13.2.2 La sélection des sujets

La sélection des sujets est une étape importante d'une étude expérimentale. Idéalement, le choix des sujets s'effectue à l'intérieur de la population à l'étude où l'on procède à une sélection aléatoire. On parle alors d'un échantillon «aléatoire». Cette technique permet généralement d'obtenir un groupe représentatif de cette population, ce qui réduit au minimum les biais de sélection.

EXEMPLE 13.5

Une étude de prévalence a révélé que, parmi les 1 800 habitants d'Anoukville, 75 avaient au moins une verrue plantaire. De ce nombre, 30 ont été choisis aléatoirement pour participer à une étude portant sur l'efficacité de l'homéopathie dans le traitement de la verrue plantaire.

En pratique, la sélection aléatoire est rarement utilisée parce que les chercheurs sont dans l'impossibilité de définir ou de connaître la population à l'étude. Pour ce faire, il

faudrait réaliser au préalable une étude de prévalence, ce qui entraînerait des coûts élevés et demanderait un temps considérable. Le recours à des sujets volontaires devient dans ce cas une solution acceptable. On parle alors d'un échantillon «électif» ou «de convenance».

Les sujets volontaires peuvent être recrutés à partir d'une population accessible, comme les étudiants d'une école ou les individus qui se présentent à un hôpital. Dans de telles circonstances, il faut comprendre que, puisqu'ils se portent volontaires, il se peut que les sujets ne soient pas représentatifs de l'ensemble de la population cible. En effet, il semble que les sujets volontaires soient généralement plus motivés et en meilleure santé que les sujets qui choisissent de ne pas participer à l'étude; leur niveau de scolarité serait également plus élevé. Bien que cela ait un effet sur la validité externe, les mesures obtenues à partir d'un échantillon électif n'auront pas pour autant un impact sur la validité interne de l'étude. Par ailleurs, il est possible aussi que les sujets d'un échantillon électif soient représentatifs de la population cible.

EXEMPLE 13.6

- Les 3 500 élèves de l'école Madeleine ont été invités à participer à une étude ayant pour but de comparer le traitement homéopathique de la verrue plantaire avec un placebo. L'invitation a été lancée par l'entremise du journal, de la radio et du personnel du service de santé de l'établissement qui informait les élèves en consultation de l'existence du projet. Comme la littérature à ce sujet rapporte une prévalence de 5 à 33 %, on estime à environ 700 le nombre d'élèves porteurs d'une verrue. Cependant, seulement 40 élèves ont manifesté leur désir de participer à l'étude. Il est donc peu probable que cet échantillon électif soit représentatif de la population à l'étude.
- Des chercheurs finlandais⁵ souhaitaient évaluer si l'administration intraveineuse de corticostéroïdes après une chirurgie cardiaque pouvait prévenir une fibrillation auriculaire postchirurgicale. Des 650 patients devant être soumis à une chirurgie cardiaque dans 3 hôpitaux, 404 ne respectaient pas les critères de sélection, et 5 ont refusé de participer à l'étude. L'échantillon électif ici est constitué de 241 patients, soit la quasi-totalité des sujets admissibles à l'étude.

Pour contrecarrer d'éventuels biais de sélection, il faut définir clairement les critères d'inclusion et d'exclusion des sujets de l'étude. Ces critères contribuent à améliorer le processus de sélection des sujets et, par conséquent, la validité interne de l'étude.

EXEMPLE 13.7

L'étude comparant le traitement homéopathique de la verrue plantaire à un placebo réalisée à l'école Madeleine comportait certains critères d'inclusion et d'exclusion. Les garçons comme les filles étaient admissibles. Les élèves qui résidaient loin de l'école se trouvaient exclus, car les visites médicales avaient lieu après les cours, à 15 h 30, heure à laquelle les élèves devaient prendre l'autobus scolaire. On avait également exclu ceux qui consommaient des médicaments ou qui étaient atteints d'une maladie chronique, ces caractéristiques pouvant influencer sur les taux de guérison.

13.2.3 La répartition des sujets par hasardisation

La **hasardisation** est une méthode rigoureuse qui consiste à répartir les sujets dans les groupes selon un procédé aléatoire, ce qui signifie que chaque sujet présente sensiblement la même probabilité de faire partie de l'un ou de l'autre groupe. Cette méthode vise à réduire non seulement les biais de sélection, mais aussi les biais de confusion. Elle permet généralement une répartition plus homogène des variables de personnes dans les groupes. Une période de *wash-out* peut parfois précéder la hasardisation et l'administration du traitement : aucun traitement actif n'est alors administré aux sujets. Cette période a pour objectif d'éliminer les effets du traitement actif habituel, susceptibles d'interagir avec ceux du traitement étudié, ou encore de s'assurer que les sujets inclus dans l'étude sont atteints de la maladie d'intérêt après l'interruption du traitement actif habituel. Ainsi, dans un essai clinique portant sur un nouveau médicament pour traiter l'hypertension artérielle, il est bon de procéder à une période de *wash-out* pour vérifier que les sujets qui seront hasardisés en fonction du traitement étudié sont effectivement hypertendus au départ.

EXEMPLE 13.8

Le fait de choisir les sujets devant recevoir un traitement A ou un traitement B selon des critères qui, à la limite, fournissent certains indices sur le traitement attribué peut conduire à une erreur systématique. Par exemple, les personnes se présentant à l'urgence pour un traumatisme oculaire bénéficieront d'un traitement de la façon suivante : les personnes qui se présentent le jour reçoivent le traitement A, et celles qui se présentent en soirée reçoivent le traitement B. Les personnes qui viennent le jour y sont peut-être incitées par des raisons particulières. Ainsi, les gens à la retraite ou qui ne travaillent pas visiteront l'urgence surtout le jour, tandis que les personnes qui travaillent le jour y viendront davantage en soirée. Par ailleurs, le fait d'avoir un traumatisme grave conduirait les personnes travaillant le jour à consulter le jour. En somme, pour plusieurs raisons, connues et inconnues, les sujets recevant le traitement A et ceux recevant le traitement B risquent d'être difficilement comparables.

On peut réaliser la hasardisation en utilisant une table de nombres aléatoires, un générateur de nombres aléatoires ou encore une autre méthode qui offre à chaque sujet la même probabilité d'être assigné à l'un ou l'autre groupe. On peut ainsi se servir d'enveloppes opaques, scellées, qui contiennent un papier indiquant le traitement à donner, ou encore appeler une centrale téléphonique qui annoncerait le traitement selon une liste de nombres aléatoires. Lorsque le recrutement s'effectue dans plusieurs milieux, il est possible de constituer des « blocs » de permutation, c'est-à-dire une petite série de traitements, choisis aléatoirement, dont une moitié serait le traitement A, et l'autre le traitement B. Ainsi, avec des blocs de six, un premier bloc pourrait être A, B, A, B, B, A, un second bloc A, A, B, B, A, B, un troisième bloc B, B, B, A, A, A, etc. Ces blocs garantissent une répartition égale des sujets dans chacun des milieux pour chaque groupe de sujets à chaque nombre donné (ici six) de sujets inclus dans l'étude.

EXEMPLE 13.9

Worrall⁶ a réalisé une étude portant sur l'utilité d'une technique stérile stricte au cours du traitement de plaies cutanées superficielles. Les sujets admissibles à l'étude ont été affectés au groupe A (technique stérile stricte) ou au groupe B (technique dite « eau et savon ») par tirage au sort. Worrall a déposé dans un sac de papier opaque six billes, soit trois billes jaunes et trois billes rouges. La sélection d'une bille jaune indiquait que le patient allait faire partie du groupe A et celle d'une bille rouge qu'il serait affecté au groupe B.

13.2.4 Les considérations éthiques particulières aux essais cliniques

De façon générale, un projet d'expérimentation sur des sujets humains n'est pas considéré comme éthiquement acceptable s'il existe des preuves irréfutables que l'un des traitements à l'étude est plus efficace que l'autre. De plus, le traitement étudié doit avoir franchi les différentes étapes qui permettent d'estimer qu'il a toutes les chances d'être plus efficace que le traitement habituel⁷.

Afin de tenter de réduire au minimum les risques de préjudices pour les sujets, les comités d'éthique des hôpitaux et des universités établissent des règles qui doivent permettre de critiquer les projets d'études avant qu'ils ne débutent.

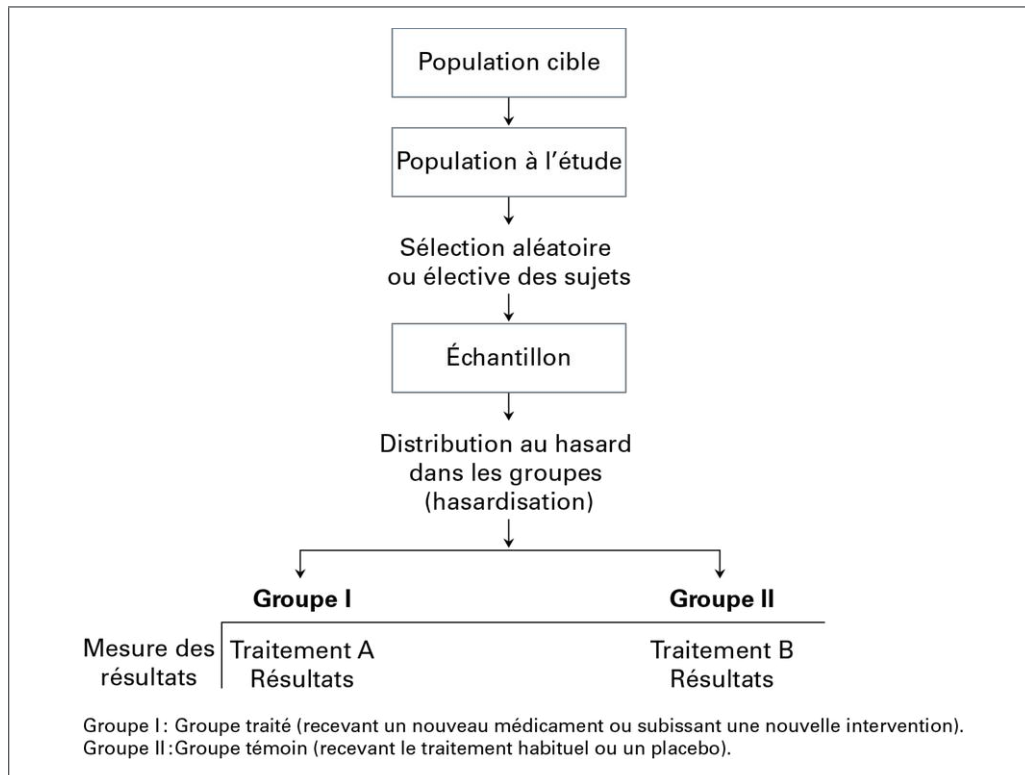
Un sujet invité à participer à une étude expérimentale doit signer, en toute connaissance de cause, un formulaire de consentement. Ce dernier contient toute l'information utile que le sujet doit connaître avant de participer au projet. Le chercheur, quant à lui, doit faire la démonstration qu'il respectera un ensemble de règles. Les aspects éthiques sont considérés à toutes les étapes de l'étude, depuis sa planification jusqu'à sa réalisation.

La qualité de l'étude, tant en ce qui a trait à sa puissance statistique qu'au plan adopté ou qu'à la rigueur du déroulement, doit nourrir la réflexion du chercheur et celle du comité d'éthique qui analyse le projet. Il serait raisonnable de s'interroger sur la pertinence de réaliser une étude dont la puissance statistique s'avérerait insuffisante, qui utiliserait une méthode de collecte d'information non rigoureuse ou à l'intérieur de laquelle le contrôle des biais de confusion serait escamoté.

13.2.5 Deux plans d'études expérimentales: en parallèle et en croisé

On utilise fréquemment deux plans d'études expérimentales hasardisées, soit le plan d'étude en parallèle et le plan d'étude en croisé.

Le **plan d'étude en parallèle** consiste à comparer deux ou plusieurs groupes simultanément. Après la hasardisation, les sujets demeurent dans le même groupe et reçoivent le même traitement tout au long de l'étude. La figure 13.1 présente de façon schématique les étapes d'une étude expérimentale menée en parallèle.

FIGURE 13.1 ■ Les étapes d'une étude expérimentale menée en parallèle**EXEMPLE 13.10**

En 1981, un groupe norvégien a publié les résultats d'une étude⁸ portant sur l'efficacité du timolol, un médicament antihypertenseur administré aux survivants d'un infarctus aigu pour réduire le risque d'un nouvel infarctus aigu du myocarde. Les sujets ont été sélectionnés à partir des admissions pour infarctus dans 20 hôpitaux de Norvège. Au total, 11 125 sujets étaient admissibles. Après l'application des critères d'inclusion et d'exclusion, 1 884 sujets ont été répartis par hasardisation dans le groupe témoin (939) ou dans le groupe traité avec le timolol (945). Tout au long de l'étude, les sujets de chaque groupe ont toujours reçu le même médicament.

Dans le **plan d'étude en croisé**, les sujets de l'étude reçoivent les deux traitements en alternance. La séquence d'administration des traitements est différente pour les deux groupes. Ainsi, le premier groupe recevra le traitement A puis le traitement B, tandis qu'on administrera le traitement B puis le traitement A au second groupe.

Ce plan d'étude est surtout utilisé lorsque les mesures varient beaucoup chez le sujet et entre les sujets. Le plan d'étude en croisé est également employé pour étudier des phénomènes plus subjectifs tels que la douleur ou le bourdonnement d'oreille.

La figure 13.2 présente les étapes d'une étude expérimentale menée en croisé.

À retenir

Plan d'étude en parallèle	Plan d'étude en croisé
Avantages	
<ul style="list-style-type: none"> • Permet un meilleur contrôle des données manquantes. 	<ul style="list-style-type: none"> • Compare le sujet à lui-même. • Exige un moins grand nombre de sujets que le plan d'étude en parallèle. • Est moins coûteux que le plan d'étude en parallèle.
Désavantages	
<ul style="list-style-type: none"> • Compare des individus potentiellement différents dans leur façon de réagir au traitement et arrivés à des stades différents de la maladie. • Exige un plus grand nombre de sujets que le plan d'étude en croisé. • Est plus coûteux que le plan d'étude en croisé. 	<ul style="list-style-type: none"> • Exige une analyse statistique complexe. • Fournit des résultats fortement influencés par les sujets perdus de vue.

13.3 L'analyse critique d'une étude expérimentale

La publication scientifique résume le travail accompli par les chercheurs et présente les grandes étapes de l'étude réalisée. Elle met en évidence l'objectif de cette étude, les méthodes utilisées pour atteindre cet objectif, les principaux résultats obtenus et propose une discussion sur le travail accompli.

De nombreux auteurs se sont intéressés à la présentation et à la critique des études expérimentales. L'énoncé CONSORT propose des normes relatives à l'organisation de la rédaction des études expérimentales¹⁰. D'autres auteurs ont suggéré des grilles de critères d'évaluation permettant de juger des résultats des études expérimentales. La médecine fondée sur les preuves ou EBM pour *Evidence-based medicine* de Guyatt et ses collègues¹¹, comprend des techniques utilisées pour guider la lecture critique de la littérature médicale. Selon cette approche, trois aspects doivent être vérifiés au moment de la critique d'un article scientifique.

1. **La validité des résultats.** Il faut déterminer si les résultats rapportés dans l'article représentent une estimation non biaisée de l'intervention (agent pharmaceutique ou autre). Il est possible de faire cette vérification à partir des informations rapportées principalement dans la section méthodologique et dans la section discussion de l'article.

- 2. L'ampleur de l'effet.** Il faut évaluer la taille ou la force de l'effet de l'intervention par rapport au traitement témoin, au moyen du calcul de la réduction du risque relatif (RRR), de la réduction du risque absolu (RRA, qui correspond à la différence des risques absolus) ou d'autres indices comme le nombre de patients à traiter (NPT) (ou en anglais NNT pour *number needed to treat*). Le NPT correspond au nombre de patients qu'il faut traiter dans l'étude pour prévenir 1 événement (le résultat de santé). Il se calcule en divisant la valeur 1 par la RRA et s'exprime en unités. La valeur des limites inférieure et supérieure de l'intervalle de confiance à 95 % entourant ces diverses mesures permettra de juger de la précision de l'estimation moyenne calculée. Plus l'intervalle est étroit, meilleure sera l'idée de l'effet réel de l'intervention dans la population.

EXEMPLE 13.12

On veut calculer le NPT d'une étude qui compare un nouveau médicament avec un placebo dans le risque de mortalité cardiovasculaire. Le risque de base (ou risque absolu chez le groupe témoin) est de 30 % : sur 100 patients, 30 seront atteints de la maladie. Le risque absolu chez le groupe expérimental est de 10 %. La RRA obtenue grâce au traitement expérimental est par conséquent de 20 % (30 % - 10 %). Le NPT correspond à $1/RRA$ ou $1/0,20$, ce qui donne 5. Il faut donc traiter cinq patients pour prévenir un cas de mortalité cardiovasculaire. Cette information est intéressante, mais il faut aussi considérer la durée de traitement pour observer ce résultat. Cela prendra-t-il deux mois ou deux ans de traitement assidu pour prévenir l'événement ? Ainsi, le NPT est interprétable uniquement si l'on connaît la durée de traitement impliquée.

- 3. La généralisation des résultats.** On fait référence ici à la validité externe de l'étude. Trois aspects doivent être examinés : les caractéristiques des sujets inclus dans l'étude, la considération de toutes les variables potentiellement confondantes et les autres **co-interventions** en cours d'étude, ainsi que la supériorité des bénéfices par rapport aux effets indésirables.

Une démarche d'analyse critique d'un article portant sur une étude expérimentale est présentée dans l'encadré 13.1. Cette grille d'évaluation est basée en partie sur les critères de Guyatt et ses collègues¹¹.

ENCADRÉ 13.1 ■ La grille d'analyse critique d'un article portant sur une étude expérimentale

Résumé

- Le résumé est-il structuré ?
- Le résumé présente-t-il le but de l'étude ?
- Y décrit-on brièvement les méthodes utilisées ?
- Y fournit-on les principaux résultats ?
- Y expose-t-on les principales conclusions ?

ENCADRÉ 13.1 ■ La grille d'analyse critique d'un article portant sur une étude expérimentale (suite)**Introduction**

- La revue des connaissances actuelles sur le sujet est-elle présentée? Les références sont-elles fournies?
- Le ou les objectifs de l'étude sont-ils adéquatement décrits?

Méthode**Devis**

- La justification du modèle d'étude est-elle présentée?

Population (ces renseignements aident à juger de la prévention des biais de sélection et de confusion)

- La population à l'étude est-elle bien définie?
- La population cible est-elle définie?
- Décrit-on le mode de sélection des sujets?
- Des moyens sont-ils mis en place pour réduire le nombre des sujets perdus de vue, les non-réponses à la sollicitation ou les refus de participer?
- Les critères d'inclusion et d'exclusion des sujets sont-ils mentionnés et permettent-ils de bien cerner la population à l'étude?
- Les lieux où les sujets sont recrutés sont-ils mentionnés?
- Les groupes sont-ils comparables?

Mesure de l'exposition et de l'effet (ces informations aident à juger de la prévention des biais d'information)

- La définition de ce que constitue un cas ou un événement est-elle claire?
- La définition de ce que constitue une exposition est-elle claire?
- La méthode de collecte de l'information est-elle bien décrite?
- L'information sur l'exposition et la maladie est-elle obtenue de la même manière pour tous les sujets? Est-elle obtenue en aveugle?
- Les caractéristiques des instruments utilisés sont-elles précisées et indique-t-on les limites de ces instruments?
- L'homogénéité de la formation des évaluateurs est-elle mesurée?

Calcul de la taille de l'échantillon

- Le calcul effectué pour fixer la taille de l'échantillon est-il précisé et expliqué?
- L'erreur de type I et la puissance statistique de l'étude sont-elles précisées?

Plan d'analyse

- Le plan d'analyse est-il présenté?



ENCADRÉ 13.1 ■ La grille d'analyse critique d'un article portant sur une étude expérimentale (*suite*)

- A-t-on pris les mesures nécessaires pour en assurer la précision ?
- Y a-t-il des sujets perdus de vue ? Si oui, sont-ils pris en considération dans l'analyse ?

Résultats

- Présente-t-on un schéma permettant d'apprécier le nombre de refus de participer et le nombre de perdus au suivi ?
- Les périodes de recrutement sont-elles clairement précisées ?
- Présente-t-on des données sociogéographiques de base pour chacun des groupes à l'étude ?
- Le résultat principal est-il clairement défini ? Quelle est l'ampleur ou la taille de son effet et quelle est sa précision ?
- A-t-on tenu compte de tous les facteurs confondants ou modifiants reconnus lors de l'analyse par stratification ou modélisation ?

Discussion et conclusion

- Les auteurs avancent-ils une interprétation des résultats principaux ?
- Ces résultats sont-ils mis en lien avec le reste de la littérature ?
- Les auteurs discutent-ils de la validité interne de l'étude, c'est-à-dire des différents biais (de sélection, d'information, de confusion) qui ont pu l'entacher ?
- Les conclusions de l'étude tiennent-elles compte de la signification clinique des résultats ?
- Les auteurs discutent-ils de la validité externe de l'étude ?

Afin de pouvoir indiquer à sa grand-mère quel traitement semble le plus efficace pour traiter sa verrue plantaire, Lisa s'est penchée sur un article intitulé *Traitement homéopathique de la verrue plantaire*¹². Cette étude servira d'exemple pour la suite du présent chapitre ; à l'occasion, d'autres exemples serviront à préciser le propos.

Le résumé de l'article portant sur le traitement homéopathique de la verrue plantaire se lit comme suit.

Objectif : Évaluer l'efficacité d'un traitement homéopathique de la verrue plantaire.

Plan de l'étude : Essai clinique hasardisé, en double aveugle et avec placebo.

Contexte : Unité de médecine familiale en milieu hospitalier.

Participants : 853 personnes ont été évaluées à l'unité de médecine familiale de décembre 1987 à janvier 1989. De ce nombre, 174 étaient admissibles (elles étaient âgées de 6 ans à 59 ans et avaient eu une ou plusieurs verrues plantaires non traitées pendant les 3 mois précédant l'étude) et avaient consenti à participer à l'étude. Pendant toute la période de l'étude (18 semaines), 162 sujets (93 %) ont été suivis.

Interventions : Un traitement homéopathique constitué de thuya 30 « centésimal hahnemannien » (CH) (un tube de 200 granules par semaine), d'antimonium

crudum 7 CH (5 granules par jour) et de nitricum acidum 7 CH (un tube de 200 granules par jour) a été administré pendant un total de six semaines. L'apparence et le goût des granules placebos étaient identiques à ceux des granules homéopathiques.

Mesure des principaux résultats : La proportion de patients guéris à l'intérieur de la période de l'étude a été établie. Un sujet était considéré comme guéri lorsque toutes ses verrues avaient disparu.

Principaux résultats : La proportion de personnes guéries à 6, 12 et 18 semaines était respectivement de 4,8 %, 13,4 % et 20 % dans le groupe homéopathie, et de 4,6 %, 13,1 % et 24,4 % dans le groupe placebo.

Conclusion : La médication homéopathique étudiée n'est pas plus efficace qu'un placebo pour traiter les verrues plantaires.

La lecture de l'article en confirme l'intérêt, et Lisa décide de poursuivre sa lecture.

13.3.1 L'introduction

L'introduction explique brièvement le cheminement du chercheur, à la lumière de son questionnement de départ. Elle permet de vérifier à quelles interrogations répond la recension des écrits présentée dans cette section et d'évaluer les propos des auteurs sur la pertinence de l'étude.

L'auteur d'un article scientifique, et à plus forte raison celui d'une étude expérimentale, doit préciser les caractéristiques de la population à l'étude. En effet, l'étude expérimentale exige un contrôle rigoureux des variables de personnes, de temps et de lieux qui peuvent influencer directement ou indirectement sur les résultats. Par ailleurs, le chercheur doit également décrire la population à laquelle il veut généraliser les résultats de son étude et préciser les caractéristiques de la population d'où provient l'échantillon qu'il a choisi. Dès lors, le lecteur est en mesure de juger de la pertinence du choix de l'échantillon.

Dans l'article qu'elle lit, Lisa repère le passage suivant :

L'étude a été réalisée à l'Unité de médecine familiale du Centre hospitalier de l'Université Laval à Sainte-Foy [aujourd'hui le CHU de Québec]. Les patients recrutés l'ont été à partir des consultations régulières à l'unité, par le réseau de télévision locale, par la radio, les journaux ou par des contacts personnels.

Les auteurs de cette étude ont choisi des volontaires (échantillon électif). Ce choix était pertinent, car les auteurs ne connaissaient pas la population à l'étude. Aucune étude de prévalence n'ayant été réalisée, une sélection aléatoire s'avérait impossible.

Les objectifs de l'étude

Les objectifs de l'étude sont habituellement présentés à la fin de la section « Introduction ». Des objectifs décrits adéquatement et suffisamment précis permettent au lecteur de suivre le raisonnement des auteurs. Les méthodes utilisées – décrites en détail dans la section « Méthode » – doivent permettre d'atteindre les objectifs.

Dans l'étude sur la verrue plantaire, on peut lire : « La présente étude a pour objectif d'évaluer l'efficacité d'un traitement homéopathe de la verrue plantaire. » Comme ils le précisent dans la section « Résumé », afin d'atteindre cet objectif, les auteurs ont opté pour une étude hasardisée et à double insu ; ils ont donc choisi le devis d'étude le plus rigoureux. Selon les informations données, le plan d'étude serait en parallèle.

EXEMPLE 13.13

Lisa se rappelle avoir lu ceci dans un autre article : « L'objectif de la présente étude est de comparer l'efficacité d'une injection de méthylprednisolone avec celle d'une injection de sérum physiologique dans l'articulation du genou de patients consultant pour la première fois pour une crise de goutte aiguë. »

On précise parfois dans l'introduction le ou les critères de jugement retenus par les auteurs, mais il arrive aussi que cet élément figure dans la section « Méthode ». Le ou les critères de jugement renvoient à la mesure de l'efficacité du traitement évalué.

Ainsi, dans l'étude portant sur la verrue plantaire, les auteurs précisent ce qui suit dans la section « Méthode » : « La verrue était considérée comme disparue si les plis cutanés étaient réapparus dans le site de la lésion. »

EXEMPLE 13.14

Une étude¹³ réalisée au Nicaragua, dont le but était d'évaluer l'efficacité d'un traitement homéopathe de la diarrhée, était hasardisée. Les auteurs ont considéré le nombre de selles par jour comme un critère de jugement de l'efficacité du traitement.

13.3.2 La méthode et les résultats

La section « Méthode » présente le plan de travail utilisé pour répondre à la question de départ. Une étude expérimentale étant d'abord et avant tout une étude comparative, le lecteur vigilant porte une attention particulière à l'identification des groupes comparés, au plan d'étude utilisé et au traitement offert au groupe témoin (placebo ou autre). De plus, il note si l'étude est hasardisée et réalisée en aveugle.

La section « Résultats » permet de juger si les travaux ont été réalisés selon le plan prévu et si des contretemps ont perturbé le déroulement des opérations (sujets perdus de vue, non-réponses à la sollicitation ou refus de participer). Pour étayer son jugement critique, Lisa cherche dans ces sections les éléments qui lui permettent de juger de l'effet du traitement, ainsi que de la validité et de la précision de l'estimation de cet effet.

L'évaluation de la validité interne de l'étude

La validité interne d'une étude consiste en sa capacité de rendre compte d'un phénomène ou d'en mesurer la valeur réelle. Une mesure valide est exempte de biais de sélection, d'information et de confusion. L'étude expérimentale permet généralement d'observer une validité interne très élevée.

Ainsi, dans l'étude portant sur la verrue plantaire, les auteurs ont choisi des sujets volontaires et ont appliqué des critères d'inclusion et d'exclusion. De plus, ils

ont réparti les sujets dans les groupes par hasardisation. Des critères précis servaient de référence pour poser le diagnostic et estimer la disparition de la verrue. Finalement, des mesures ont été prises pour réduire le nombre de sujets perdus de vue. Toutes ces mesures ont été mises en place pour réduire les biais.

La prévention et le contrôle des biais de sélection

La sélection des sujets est une étape importante de la réalisation d'une étude expérimentale. Une description détaillée du processus de sélection doit par conséquent être fournie.

Il est préférable d'exclure certains sujets d'une étude. Ainsi, on n'admettra pas les femmes enceintes parce que, dans leur cas, l'innocuité du traitement n'a pas été démontrée. On peut aussi décider que les sujets de l'étude appartiendront à certains groupes d'âge ou qu'ils ne devront pas souffrir d'une maladie chronique. Ces exclusions précédant la hasardisation ont l'avantage de favoriser l'homogénéité des groupes. En revanche, une sélection trop rigide limitera les conclusions de l'étude à un groupe très restreint d'individus, ce qui réduira la validité externe.

Labrecque et ses collègues ont indiqué ceci dans leur article :

Tous les sujets de 6 ans à 59 ans ayant au moins une verrue plantaire étaient admissibles. Les critères d'exclusion comprenaient : avoir été traité pour une verrue plantaire dans les trois mois précédant l'étude, avoir une verrue en mosaïque, prendre un médicament prescrit, être atteint d'une maladie chronique ou d'une immunodéficience. Les femmes enceintes et celles qui allaitaient n'étaient pas admissibles, de même que les patients demeurant à l'extérieur de la ville de Québec.

Les auteurs de l'étude ont donc choisi d'exclure certains sujets lors de la sélection. Leurs critères d'exclusion semblent réalistes. En effet, ils ont choisi d'exclure les sujets résidant en dehors des limites de Québec, car l'observance du traitement aurait pu en être perturbée; ils ont tenu compte de facteurs pouvant avoir un effet sur la guérison des verrues tels que l'âge, les maladies chroniques et le fait d'être enceinte ou d'allaiter, et ils ont exclu les cas de diagnostic incertain. Le tableau 13.1, à la page suivante, tiré de l'article, renseigne le lecteur sur les motifs d'exclusion.

Il est possible de prévenir les biais de sélection à l'étape de l'échantillonnage. Le chercheur peut alors s'assurer de la qualité des banques de données ou des listes à partir desquelles il sélectionnera les cas; il prend les moyens pour éviter de perdre de vue des sujets au cours de l'étude et pour limiter le nombre de refus d'y participer. Les résultats d'un essai clinique dépassant 20 % d'abandons ou de pertes au suivi témoignent souvent de la présence d'un biais de sélection¹⁴.

À l'étape de l'analyse, le contrôle des biais de sélection est presque impossible. Le chercheur doit tenter d'estimer l'impact d'un éventuel biais de sélection sur les résultats obtenus en tenant compte dans son analyse des refus de participer à l'étude, ainsi que des sujets perdus de vue et des abandons en cours d'étude. Pour ces deux derniers cas, le chercheur peut effectuer une **analyse de sensibilité**. En l'absence d'information sur le devenir réel des sujets perdus de vue ou ayant abandonné en

cours d'étude, le chercheur refait alors les analyses en attribuant à chacun de ces sujets d'abord le moins favorable des effets, puis le plus favorable d'entre eux, ou d'autres scénarios pour juger si les nouvelles estimations sont similaires aux principaux résultats observés.

TABLEAU 13.1 ■ Les motifs d'exclusion des sujets lors du recrutement

Motif	Nombre (%) de patients (<i>n</i> = 853)
Exclus après un questionnaire téléphonique	485 (56,9)
Verrue traitée au cours des trois derniers mois	127 (26,2)
Patient désirant de l'information seulement	123 (25,4)
Résidence hors des limites de Québec	69 (14,2)
Autres lésions plantaires	67 (13,8)
Maladies chroniques traitées	47 (9,7)
Verrue non plantaire	24 (4,9)
Âge < 6 ans ou > 59 ans	21 (4,3)
Patiente enceinte ou qui allaite	7 (1,4)
Exclus après un examen physique	111 (13,0)
Autres lésions plantaires	69 (62,2)
Absence de <i>microthrombi</i>	24 (21,6)
Verrue en mosaïque	18 (16,2)
Refus de participer	83 (9,7)

EXEMPLE 13.15

Le groupe norvégien¹⁵ qui a étudié l'efficacité du timolol en le comparant à un placebo s'est trouvé dans une situation où certains sujets ont cessé de suivre le traitement comme il était recommandé à cause d'effets indésirables; d'autres sujets ont abandonné le traitement en cours de route. Enfin, d'autres ont été simplement perdus de vue. Cette erreur systématique a été en quelque sorte « corrigée » en procédant à une analyse de sensibilité. Les différences observées étaient assez similaires, que les sujets ayant abandonné l'étude ou ayant été perdus de vue aient été inclus ou exclus des calculs. Les auteurs ont conclu que le timolol était efficace pour réduire la mortalité postinfarctus.

Bien qu'il existe des moyens de réduire au minimum les effets attribuables aux sujets perdus de vue, ce phénomène constitue un problème majeur que l'analyse

statistique la plus rigoureuse ne peut résoudre. Les propos des auteurs quant à l'impact de ces sujets sur les résultats obtenus sont très importants.

Dans l'article dont Lisa fait l'analyse critique, on lit ce qui suit dans la section « Discussion » :

Nous avons pris soin de minimiser les sources de biais. La proportion des « perdus de vue », qui peut entraîner un biais de sélection dans un essai clinique, a été réduite à un minimum (7 % dans chaque groupe). Une nouvelle analyse des données ainsi que l'attribution du meilleur résultat possible (disparition de la verrue) à ceux qui avaient abandonné le groupe traité par homéopathie et du pire résultat (verruve toujours présente) à ceux qui avaient quitté le groupe traité par un placebo n'a pas changé les résultats.

Le lecteur vigilant doit donc s'interroger sur les aspects suivants : les patients ont-ils été suivis jusqu'à la fin de l'étude ? Sinon, la situation est-elle, toute proportion gardée, comparable entre les groupes ? Les auteurs en tiennent-ils compte dans l'analyse ?

La prévention des biais d'information

Comme on le sait, le sujet, l'intervenant qui effectue l'évaluation du sujet ou le chercheur qui est informé de la nature du traitement administré sont influencés par cette connaissance. En effet, le sujet qui sait qu'il a consommé un placebo pourra difficilement croire aux vertus curatives de ce traitement. En revanche, s'il ignore ce qu'il consomme, il peut attribuer des propriétés au produit même s'il s'agit d'un placebo, et sa croyance en l'efficacité du médicament est peut-être ce qui le soulage. Il en va de même pour le chercheur qui croit à l'efficacité d'un traitement. Son évaluation du résultat sera influencée par le préjugé qu'il risque de nourrir s'il sait qui reçoit le placebo et qui reçoit le traitement.

La technique en aveugle permet de prévenir le biais d'information. La technique en double ou triple aveugle (*voir le chapitre 6*) est préférable à la technique en simple aveugle, et cette dernière vaut mieux qu'aucune technique. Il revient au lecteur de juger s'il se trouve vraiment devant une étude menée en aveugle : certaines sont prétendument effectuées de cette façon, alors que ce n'est pas le cas.

EXEMPLE 13.16

Une compagnie a subventionné la réalisation d'une étude comparant un nouvel antibiotique à un placebo. Les sujets sélectionnés étaient des volontaires, et la distribution dans les groupes était hasardisée. De plus, l'étude était dite en double aveugle. Or, le nouvel antibiotique se composait d'une molécule de la famille des érythromycines, des médicaments qui provoquent des troubles gastro-intestinaux. Ainsi, l'antibiotique à l'étude a provoqué des nausées et des crampes abdominales chez 66 % des sujets, alors que ces symptômes se sont manifestés chez seulement 8 % des sujets recevant le placebo. Il était donc possible pour les sujets et pour les évaluateurs de présumer qu'en présence de troubles gastro-intestinaux, il y avait de plus fortes chances d'appartenir au groupe traité avec l'antibiotique. Par conséquent, l'étude n'était pas véritablement menée en aveugle.

Outre la technique en aveugle, le chercheur doit préciser les mesures prises pour réduire ou éviter les biais d'information. Ainsi, la méthode de collecte d'information doit être bien décrite. Il faut s'assurer de la qualité des instruments utilisés, qu'il s'agisse d'un questionnaire ou d'un examen diagnostique. De plus, le chercheur doit mentionner les caractéristiques et les limites des instruments utilisés, mais faute d'espace, il peut aussi référer le lecteur à un article de validation pour justifier le choix de ses instruments. Lorsqu'il y a plus d'un évaluateur pour la prise d'une exposition d'intérêt, l'homogénéité de leur formation doit être assurée. Cette homogénéité est souvent démontrée par des résultats sur l'accord interjuges (coefficient de corrélation ou mesure Kappa).

Dans l'étude portant sur le traitement de la verrue plantaire, les auteurs ont demandé à des médecins d'évaluer le diagnostic posé et la guérison obtenue :

Recourir à plusieurs observateurs peut entraîner une erreur de classification à l'étape de la sélection des patients et à celle de l'évaluation du résultat. Afin de réduire cette erreur, des critères objectifs ont été utilisés. Le diagnostic de verrue plantaire était posé lorsque les trois signes suivants étaient présents à l'examen : l'hyperkératose, la disparition des plis cutanés et la présence de *microthrombi*. La verrue était considérée comme disparue si les plis cutanés étaient réapparus dans le site de la lésion. Un patient était considéré comme guéri si toutes les verrues étaient disparues. Une étude de reproductibilité inter-observateurs du diagnostic de verrue plantaire et de l'évaluation de la guérison de la verrue a été réalisée préalablement à la présente étude. En raison d'un accord inter-observateurs modéré (80 %, $K = 0,6$), les critères d'évaluation ont fait l'objet de discussions au sein de l'équipe, et des exemples illustrés ont été présentés au cours de la formation qui a précédé l'étude. Afin d'accroître la spécificité du diagnostic clinique, deux médecins devaient poser le diagnostic de verrue plantaire.

On voit donc que les auteurs ont mis en œuvre une série de moyens pour réduire les biais d'information et que ces moyens semblent appropriés.

À l'étape de l'analyse, un contrôle des biais d'information est généralement très difficile si le chercheur ne connaît pas les valeurs de sensibilité ou de spécificité des instruments utilisés. En l'absence de cette information, il doit tenter d'estimer les effets de tels biais en effectuant des mesures indirectes ou des comparaisons multiples ; il peut par exemple recourir à différents instruments pour mesurer le même phénomène. Toutefois, ces méthodes ne sont que rarement utilisées, et il est préférable de privilégier la prévention de ce type de biais.

La prévention des biais de confusion

Au moment de l'élaboration du protocole, la répartition des sujets dans les groupes par hasardisation demeure le critère le plus important à respecter si l'on veut éviter les biais de confusion. Le risque d'une erreur systématique doit toujours être envisagé en l'absence de hasardisation. Si elle est réussie, la hasardisation assure l'homogénéité et la comparabilité des groupes tant pour les facteurs de confusion connus que pour ceux encore inconnus. Par ailleurs, l'analyse en intention de traitement consiste à analyser les sujets selon l'allocation initiale du sujet à un traitement, sans tenir compte de l'observance ou de la fidélité du sujet au traitement. Aucun ajustement des résultats n'est alors nécessaire. Cette technique assure le respect de la distribution de tous les facteurs

confondants connus et inconnus entre les groupes. De ce fait, elle permet également d'obtenir l'efficacité réelle d'un traitement. En contrepartie, il est possible d'effectuer une analyse par protocole, c'est-à-dire d'analyser les sujets selon le traitement réellement reçu en cours d'étude. Les résultats de ce type d'analyse devraient alors être ajustés en fonction des variables confondantes puisqu'il y aurait non-respect de la hasardisation. Une analyse par protocole permettrait de mesurer l'efficacité théorique d'un traitement.

Dans l'étude portant sur la verrue plantaire, les auteurs présentent les caractéristiques de la population étudiée. Le tableau 13.2 résume certaines des informations présentées dans le tableau original.

TABLEAU 13.2 ■ Les caractéristiques des sujets selon le type de traitement

Caractéristique	Type de traitement	
	Homéopathique <i>n</i> = 88 nombre (%)	Placebo <i>n</i> = 88 nombre (%)
Âge < 20 ans	17 (19,8)	33 (37,5)
Sexe féminin	54 (62,8)	59 (67,0)
Verrue présente depuis moins de un an	41 (47,7)	59 (67,0)
Verrue unique	36 (41,9)	38 (43,2)

La hasardisation semble avoir été efficace en ce qui a trait à certaines variables, mais moins efficace quant à certaines autres. Les auteurs soupçonnent qu'il existe une différence entre les groupes, notamment en ce qui concerne la proportion des sujets âgés de moins de 20 ans et l'âge de la verrue. Les auteurs précisent dans la section « Résultats » qu'une « analyse stratifiée a été réalisée pour les variables possiblement associées avec la guérison des verrues plantaires : l'âge, le sexe, le nombre de verrues, la douleur, l'âge de la verrue... » En outre, ils indiquent :

[...] les résultats sont demeurés inchangés pour toutes les variables à l'exception de l'âge. Chez les participants de moins de 20 ans, nous avons observé une différence clinique, mais non statistiquement significative, en faveur du traitement homéopathique à la sixième semaine. Cette différence disparaît toutefois avec le temps et, à 18 semaines, elle n'était plus significative, ni cliniquement, ni statistiquement.

Il ne semble donc pas y avoir de biais de confusion assez important pour que les résultats soient invalidés.

L'âge est une variable confondante bien connue, et sans doute la plus fréquente. Une étude réalisée à partir d'un petit échantillon constitué de sujets de tous les âges peut poser des problèmes d'interprétation; il peut en effet se révéler impossible d'analyser des sous-groupes parce que leurs effectifs sont trop petits. Il est alors préférable de restreindre la participation à l'étude à certains groupes d'âge. Lorsque les variables

confondantes sont bien connues, le chercheur décide dès l'étape de l'échantillonnage s'il entend faire une analyse par stratification (**analyse par sous-groupes**), par restriction ou par assortiment. Ainsi, dans l'étude portant sur la verrue plantaire, les auteurs ont choisi d'exclure les enfants âgés de moins de cinq ans au moment de la sélection. Il s'agit de la technique par restriction.

Certains facteurs, s'ils ne sont pas considérés dans l'analyse, peuvent être à l'origine de confusion. Ces facteurs, qui se manifestent après la hasardisation, sont les co-interventions et l'observance du traitement.

EXEMPLE 13.17

Un sujet a reçu aléatoirement le traitement A pour soigner sa verrue plantaire. En cours d'étude, il décide de consulter un médecin pour que celui-ci fasse disparaître sa verrue avec de l'azote liquide. Pour être bien sûr d'obtenir les résultats escomptés, il applique également sur sa verrue des disques et d'autres préparations à base d'acide lactique et d'acide salicylique. Tous ces traitements ont un effet réel ou présumé et peuvent par conséquent modifier l'évolution de la verrue plantaire. L'effet du traitement A ne peut alors être évalué adéquatement.

Si une ou des co-interventions se produisent en cours d'étude, le chercheur doit obtenir du patient le maximum d'information sur celles-ci. Il doit de plus estimer les effets qu'elles peuvent avoir et en tenir compte dans l'analyse. Une analyse stratifiée peut être réalisée.

Dans l'étude de Labrecque et ses collègues, les auteurs écrivent dans la section « Résultats » :

[...] la proportion de ceux qui n'avaient pas utilisé une autre modalité de traitement pour traiter leur verrue était légèrement inférieure dans le groupe traité par homéopathie.

Le pourcentage de sujets n'ayant recouru à aucun autre traitement était de 76,3 % dans le groupe traité par homéopathie et de 85,4 % dans le groupe recevant un placebo. Les auteurs précisent plus loin qu'une analyse stratifiée a été réalisée pour la variable *utilisation d'un autre traitement* (oui *versus* non) et que les résultats sont demeurés inchangés selon ces deux strates, c'est-à-dire qu'aucune différence statistique ou clinique n'a été mise en évidence.

Le taux de guérison est souvent proportionnel à l'observance du traitement. Ainsi, les chances de guérir une cellulite de l'avant-bras sont moindres chez un patient qui consomme une seule capsule d'antibiotique que chez celui qui en prend une toutes les 8 heures pendant 10 jours et qui respecte ainsi la posologie.

En cours d'étude, le chercheur doit mesurer jusqu'à quel point les sujets observent le traitement et il doit en tenir compte dans l'analyse. Une stratification selon les taux d'observance aide le lecteur à se forger une opinion quant à l'efficacité réelle du médicament ou de l'intervention faisant l'objet de la recherche. De plus, elle lui fournit un indice sur la façon dont les patients réagiront s'ils n'observent que partiellement le traitement, phénomène qui survient fréquemment dans un contexte clinique.

EXEMPLE 13.18

Dans une étude portant sur les effets d'un nouveau médicament contre l'hypertension, une analyse stratifiée a permis de constater que ce médicament réduit significativement la pression artérielle lorsque le taux d'observance calculé chez les sujets est égal ou supérieur à 80 %. Un taux d'observance inférieur à 80 % ne permet pas de contrôler efficacement la pression artérielle.

L'évaluation de la validité externe de l'étude: la précision de l'étude

L'erreur aléatoire peut être réduite de deux façons : par l'utilisation d'instruments de mesure précis et par le recours à une taille d'échantillon suffisamment importante.

Les instruments de mesure utilisés dans une étude expérimentale peuvent être nombreux et variés, allant d'un questionnaire à un instrument de laboratoire. Chacun de ces instruments possède un degré de précision appréciable. Le chercheur qui utilise des instruments doit s'assurer de leur qualité. Une étude de validation est généralement réalisée avant le début de l'étude afin d'estimer la qualité des instruments ; en cours d'étude, d'autres tests peuvent être effectués. Dans l'article dont l'étude fait l'objet, les caractéristiques des instruments et les valeurs de la concordance inter-examineurs pour le diagnostic clinique de verrue plantaire ont été présentées dans la section « Méthode ».

Dans cette section, l'auteur précise aussi la méthode de calcul de la taille de l'échantillon et fixe l'erreur de type I consentie. Plus la taille de l'échantillon sera grande, plus la puissance statistique de l'étude le sera aussi. La puissance statistique de l'étude désigne, pour une taille d'échantillon donnée, sa capacité de déceler une différence minimale fixée lorsque cette différence existe réellement dans la population. Pour chaque type d'étude, des formules mathématiques permettent de calculer la taille appropriée de l'échantillon. Les études expérimentales atteignent généralement une puissance statistique de 80 % ou plus.

Dans l'article sur la verrue plantaire, les auteurs notent ceci dans la section « Discussion » :

Étant donné que 24 % des sujets étaient guéris après 18 semaines, la probabilité de détecter une différence statistiquement significative entre les deux groupes, en présumant un taux de guérison de 50 % dans le groupe traité par homéopathie, était de 90 %.

La puissance de l'étude était donc de 90 %. On se rappelle que les chercheurs doivent estimer la puissance statistique de l'étude au moment de sa planification. Le calcul de puissance statistique devrait normalement figurer dans la section « Méthode » de l'article.

La présentation du plan d'analyse

Dans la section « Méthode », les auteurs doivent également présenter la méthode d'analyse des données et les différents outils statistiques utilisés.

Dans l'étude portant sur la verrue plantaire, la description du plan d'analyse statistique n'est ni systématique ni très élaborée. La mesure d'accord Kappa a été utilisée

pour apprécier la concordance inter-examineurs en ce qui concerne le diagnostic. La taille de l'échantillon a été calculée à l'aide d'un logiciel, et la signification statistique des résultats est estimée par des intervalles de confiance à 95 % (c'est-à-dire seuil $\alpha = 0,05$, voir le chapitre 11).

EXEMPLE 13.19

Dans une autre étude¹⁶, on trouve une description plus rigoureuse d'un plan d'analyse.

Une analyse du khi carré, un test exact de Fisher et un test de t de Student pour échantillons indépendants ont été utilisés pour comparer les caractéristiques des sujets exposés et non exposés. Parce que toutes les données concernant l'exposition étaient concentrées du côté positif avec 45 % ou plus et classées comme «non exposé» par chacune des procédures de recueil de l'information, des méthodes non paramétriques incluant le test de rang de Wilcoxon et le test de Kruskal-Wallis ont été utilisées pour comparer les niveaux d'exposition entre les groupes. L'indice de corrélation de rang de Spearman a été utilisé pour établir la corrélation entre les variables non catégorielles. Lorsque l'exposition était catégorisée en «exposé» et «non exposé», l'accord entre les méthodes de classification de l'exposition était évalué par le Kappa. L'erreur de classification était calculée comme le nombre de sujets pour lesquels les méthodes divergeaient divisé par le nombre total de sujets. Tous les calculs ont été effectués à l'aide du logiciel SAS.

Bien sûr, il est difficile de juger de la pertinence de tous les tests statistiques utilisés par les auteurs. Cependant, lorsque ceux-ci rapportent explicitement leur plan d'analyse statistique, ils gagnent en crédibilité parce qu'il est facile de consulter, au besoin, des statisticiens ou des épidémiologistes pour juger de son à-propos.

13.3.3 La discussion et la conclusion

La discussion et la conclusion présentent la critique que font les auteurs de l'étude qu'ils ont menée.

L'évaluation de la validité interne de l'étude

Les auteurs présentent généralement les forces et les limites de leur étude dans la discussion. Ils peuvent passer en revue les biais (sélection, information et confusion) qui ont entaché l'étude ou qui auraient pu le faire. Dans chaque cas, ils expliquent les moyens utilisés pour contrôler ces biais ou, encore, ils précisent jusqu'à quel point et dans quelle direction (sous-estimation ou surestimation) les biais résiduels peuvent influencer sur les résultats obtenus.

Dans la section «Discussion», Labrecque et ses collègues ont indiqué :

Nous avons pris soin de minimiser les sources de biais. La proportion des «perdus de vue», qui peut entraîner un biais de sélection dans un essai clinique, a été réduite à un minimum (7 % dans chaque groupe). Une nouvelle analyse des données ainsi que l'attribution du meilleur résultat possible à ceux qui avaient abandonné le groupe traité par homéopathie et du pire résultat à ceux qui avaient

quitté le groupe traité par un placebo n'a pas changé les résultats. On a procédé à une analyse stratifiée des données pour détecter des biais de confusion, mais aucune variable ne semble avoir introduit ce type d'erreur. Un biais d'information (erreur de classification) était possible parce que les patients qui ne revenaient pas à l'hôpital pour une visite de contrôle rapportaient eux-mêmes la présence ou l'absence de verrues. Nous pouvons présumer que, dans le contexte d'un essai clinique en double aveugle, il doit s'agir là d'un biais non différentiel qui tend à minimiser les différences entre les deux groupes. Ce biais ne s'est probablement pas produit parce que l'analyse stratifiée a démontré des résultats similaires, quelle que soit la personne qui a reconnu l'issue du traitement; le médecin de famille ou le patient.

L'évaluation de la taille de l'effet et sa précision

Dans l'étude sur la verrue plantaire, les résultats ont montré que 24,4 % des sujets dans le groupe placebo ont obtenu une guérison contre 20,0 % des sujets dans le groupe homéopathique. Cette différence de risques (-4,4 %, ic 95 %, -17,0 à 8,5 %) correspondait à une augmentation du risque absolu, ce qui semblait favoriser le traitement placebo au détriment du traitement par homéopathie. Cependant, cette différence était non significative. Par conséquent, la taille de l'effet ne peut être considérée puisque les deux traitements ne sont pas statistiquement différents. Aucun calcul de NPT n'était pertinent.

L'évaluation de la validité externe de l'étude

L'étude expérimentale offre habituellement une validité interne élevée. L'obtention de cette forte validité interne peut se faire au détriment de la validité externe de l'étude si les critères de sélection des sujets sont trop stricts. À ce propos, il est bon de vérifier le **diagramme de flux** (ordinogramme, *flow-chart*) des sujets lors du recrutement pour savoir jusqu'à quel point les sujets inclus dans l'étude sont représentatifs de la population cible. De plus, une comparaison avec d'autres études portant sur un même sujet est souvent effectuée. Cela permet au lecteur de juger de la valeur des résultats et d'estimer si ces résultats sont généralisables ou applicables à une autre population, comme à celle de sa pratique.

Dans l'étude de Labrecque et ses collègues, les auteurs concluent :

Le traitement homéopathique s'avère aussi efficace que le traitement placebo. Le taux de guérison par le traitement homéopathique est nettement inférieur aux taux suggérés dans des études antérieures et est inférieur aux taux de succès des traitements conventionnels. Dans une série d'essais cliniques, les verrues étaient disparues après trois mois de traitement aux kératolytiques ou à l'azote liquide, ou les deux, dans des proportions allant de 50 à 80 % des patients. Bien que le traitement homéopathique utilisé dans cette étude soit indiqué pour les verrues plantaires, d'autres traitements homéopathiques ou des durées de traitement plus longues auraient pu donner de meilleurs résultats. D'autres recherches seront nécessaires afin d'évaluer l'efficacité de ce type d'intervention. Quoi qu'il en soit, nous ne considérons pas que le traitement homéopathique étudié présente un intérêt clinique pour les patients atteints de verrues plantaires.

À l'occasion d'une seconde rencontre à la polyclinique, Lisa explique à sa grand-mère : «J'ai vérifié l'efficacité de plusieurs traitements de la verrue plantaire. Il faut que vous sachiez que, même sans traitement, votre verrue pourrait disparaître. Il semble qu'environ deux verrues sur trois guérissent sans traitement en un an. Si vous désirez que la vôtre disparaisse plus rapidement, vous pouvez utiliser les médicaments offerts en vente libre dans les pharmacies. La plupart des patients guérissent en quelques mois avec ces traitements. Quant au traitement homéopathique, je doute qu'il soit efficace si je me fie à l'étude que j'ai consultée. Si j'étais à votre place, je choiserais d'appliquer le traitement que nous utilisons fréquemment ici, c'est-à-dire un mélange de vaseline et d'aspirine. Ça ne coûte pas cher et ça fonctionne!»

13.3.4 La déclaration financière

Dans le cas d'un essai clinique, il est toujours pertinent de vérifier la source du financement de l'étude. Certains journaux scientifiques demandent aux auteurs de rapporter tout contrat en lien avec l'industrie pharmaceutique. De plus, si la subvention provient de commandites de celle-ci, certains auteurs spécifient s'il s'agit d'une subvention sans restriction, c'est-à-dire que la compagnie pharmaceutique ne se réserve aucun droit de regard, peu importe les résultats obtenus. D'autres auteurs pourraient mentionner dans l'article que les analyses statistiques ont été refaites par un groupe de chercheurs indépendants. Toutes ces stratégies assurent le plus de transparence possible au moment de la réalisation de l'étude et confèrent plus de crédibilité aux résultats.

En résumé

L'étude expérimentale est le modèle d'étude idéal si l'on souhaite comparer des traitements ou des interventions, car elle s'apparente à une expérience de laboratoire. Ainsi, elle permet de réduire ou d'éviter de nombreux biais grâce à la rigueur de sa méthode. La hasardisation est sa principale caractéristique.

Avant sa mise en œuvre, toute étude faisant appel à des humains ou à du matériel biologique humain doit être évaluée par un comité d'éthique et de recherche. Il ne serait pas éthique de comparer un nouveau traitement à un placebo lorsqu'un traitement efficace existe ou d'administrer un traitement expérimental qui pourrait comporter des risques importants pour les sujets.

Pour faire une analyse critique de ce type d'étude, on doit définir le type de plan d'étude utilisé et en connaître les principales forces et faiblesses. Il faut par la suite en évaluer la validité interne en tentant de repérer les biais de sélection, d'information et de confusion et en analysant les méthodes de contrôle utilisées pour réduire les effets négatifs de ces biais sur les résultats. Enfin, on évalue la précision de l'étude en appréciant sa puissance statistique et l'on porte un jugement sur sa validité externe. L'analyse critique d'une étude expérimentale permet d'en juger les résultats et les conclusions qui, plus tard, pourront être intégrés à la pratique clinique.

Exercices**Voir la section complète des exercices sur le Web.**

1. Parmi les énoncés suivants, lequel s'applique uniquement aux études expérimentales?
 - a) Les sujets sont sélectionnés aléatoirement dans la population à l'étude.
 - b) Les chercheurs ont constitué un ou plusieurs groupes témoins.
 - c) Les chercheurs précisent les critères de jugement qu'ils ont retenus.
 - d) Les chercheurs utilisent des instruments de mesure dont les caractéristiques sont connues.
 - e) Les chercheurs forment des groupes de sujets et déterminent lequel recevra le traitement étudié.
2. Quel est le but de la hasardisation dans un essai clinique?
 - a) S'assurer que l'expérience sera menée en aveugle.
 - b) Répartir les variables de personnes de façon homogène entre les groupes.
 - c) Inclure le même nombre de sujets dans chaque groupe.
 - d) S'assurer que le groupe à l'étude est représentatif de la population générale.
 - e) Contrôler les biais d'information.
3. Quel type de biais l'utilisation d'une technique en aveugle permet-elle d'éviter dans une étude expérimentale?
 - a) Biais de sélection.
 - b) Biais d'information.
 - c) Biais de confusion.
4. Quel type de biais l'utilisation d'un plan d'étude en croisé permet-elle d'éviter dans une étude expérimentale?
 - a) Biais de sélection.
 - b) Biais d'information.
 - c) Biais de confusion.

Pour les réponses, voir le corrigé à la fin du manuel.

CHAPITRE 14

Collaboration à la 3^e édition :

Helen Trottier

Marie-Hélène Mayrand

L'analyse critique des études à visée étiologique : études non expérimentales cas-témoins et de cohorte(s)

Au terme de ce chapitre, vous pourrez :

- juger de la pertinence et de la validité scientifique d'une étude à visée étiologique.

Objectifs spécifiques	Contenu
14.1 Expliquer les particularités des études à visée étiologique.	14.1 L'élaboration des études à visée étiologique 14.1.1 Les particularités des études cas-témoins et de cohorte(s) 14.1.2 L'exemple du café et du cancer du pancréas
14.2 Procéder à l'analyse critique des études cas-témoins et de cohorte(s) à l'aide d'une grille d'évaluation. <ul style="list-style-type: none">• Plan d'étude approprié• Prévention et évaluation de l'impact des biais de sélection et d'information et contrôle du biais de confusion• Précision de l'étude• Signification statistique et signification clinique des résultats	14.2 L'analyse critique des études cas-témoins et de cohorte(s) 14.2.1 L'analyse critique d'une étude cas-témoins 14.2.2 L'analyse critique d'une étude de cohorte(s) 14.2.3 La recommandation clinique à la suite de l'analyse clinique 14.3 Les considérations éthiques





Mise en situation

Geneviève a 40 ans. Elle travaille comme secrétaire juridique dans un bureau d'avocats très réputé. Aujourd'hui, elle consulte l'omnipraticienne Lisa pour une vague douleur abdominale. Manifestement anxieuse, Geneviève explique à Lisa qu'elle a lu dans le dernier numéro de la revue *Jessica* que le café pouvait causer le cancer du pancréas. Or, elle en boit de 6 à 10 tasses par jour, et son père est décédé l'an passé d'un cancer du pancréas à l'âge de 61 ans. Il buvait énormément de café et fumait beaucoup.

Lisa questionne Geneviève et procède à un examen physique complet. Elle lui propose aussi quelques examens complémentaires. Elle lui dit qu'elle comprend ses inquiétudes. Bien qu'il existe plusieurs bonnes raisons de ne pas boire 10 tasses de café par jour, elle lui explique que le café ne serait pas un facteur de risque du cancer du pancréas. Par contre, le tabac le serait. Devant l'intérêt de Geneviève d'en savoir plus, Lisa décide de mettre à jour ses connaissances sur les facteurs de risque reconnus du cancer du pancréas. Elle lui donne rendez-vous deux semaines plus tard et lui recommande, dans l'intervalle, de modérer sa consommation de café et de recourir à des techniques de relaxation.

Pertinence clinique

Le cancer du pancréas est une maladie rare, mais il constitue l'une des principales causes de mortalité par cancer en Amérique du Nord. L'étiologie de ce cancer demeure mal connue. À ce jour, l'efficacité du traitement est limitée, et la plupart des personnes qui en sont atteintes meurent moins de un an après le diagnostic. Tout comme le cancer du pancréas, plusieurs autres maladies ne peuvent être traitées avec succès. Une connaissance accrue de leurs facteurs de risque pourrait contribuer à une meilleure prévention de la maladie.

Pour des raisons éthiques, les études expérimentales se prêtent mal à l'étude des causes des maladies. Les connaissances en ce domaine sont le plus souvent obtenues grâce à des études non expérimentales d'observation à visée étiologique : l'étude cas-témoins et l'étude de cohorte(s). La comparaison de groupes de sujets atteints ou non d'une maladie ou de groupes exposés et de groupes non exposés à certains facteurs peut contribuer à cerner le rôle que jouent des facteurs particuliers dans le développement de cette maladie. Le clinicien doit être capable de juger de la qualité des études d'observation à visée étiologique et savoir interpréter correctement leurs résultats.

De manière à pouvoir répondre adéquatement aux questions de Geneviève, Lisa réunit ses collègues pour discuter des facteurs de risque du cancer du pancréas et pour revoir les principes qui régissent les études à visée étiologique, en particulier les études cas-témoins et de cohorte(s).

Tout comme eux, nous passerons en revue les principes qui sous-tendent ces études. Nous discuterons d'une étude cas-témoins sur le lien entre le tabagisme, la consommation d'alcool et de café et le cancer du pancréas à l'aide d'un article sur le sujet intitulé *Tobacco, alcohol, and coffee and cancer of the pancreas, A population-based, case-control study in Quebec, Canada*¹. Nous présenterons également une étude de cohorte portant sur le même sujet et faisant l'objet de l'article *A cohort study of smoking, alcohol consumption, and dietary factors for pancreatic cancer (United States)*².

Danielle, l'infirmière qui participe à la rencontre, suggère à Lisa d'utiliser, comme elle l'a fait pour les études descriptives de population, la grille d'évaluation des études d'observation élaborée par l'initiative internationale STROBE³ (voir le chapitre 9).

14.1 L'élaboration des études à visée étiologique

Lorsqu'on étudie pour la première fois l'étiologie d'une maladie, il peut être difficile de déterminer les facteurs de risque à rechercher. Les connaissances en biologie, en biochimie et en physiologie, de même que les études écologiques dont on dispose, permettent de formuler les premières hypothèses. Les études écologiques permettent d'identifier des corrélations entre différents facteurs de risque et des maladies, mais uniquement au niveau populationnel. Pour cette raison, elles servent avant tout à formuler des hypothèses qui doivent par la suite être investiguées par d'autres devis qui permettront de chercher des associations au niveau «individuel». Ainsi, en ce qui concerne le cancer du pancréas, des chercheurs avaient mené une étude écologique, publiée en 1983, qui fut l'une des premières à désigner la consommation de café comme étant un facteur de risque potentiel de cancer chez les hommes⁴. Ils ont établi une corrélation entre les taux ajustés selon l'âge pour différents cancers et plusieurs variables, dont les habitudes de consommation de café dans une vingtaine de pays. Pour sa part, Spector⁵ a remarqué que, entre 1948 et 1978, l'évolution de la mortalité par cancer du pancréas suivait une courbe parallèle à celle des importations de café. Bernarde et Weiss⁶, quant à eux, ont constaté qu'aux États-Unis, il y avait eu une augmentation puis une diminution de la consommation de café, suivies, une dizaine d'années plus tard, d'une augmentation puis d'une diminution du cancer du pancréas. Bref, dans l'ensemble, les études écologiques laissaient croire à l'existence d'un lien entre la consommation de café et le cancer du pancréas. Ces études ne permettaient toutefois pas de vérifier si les personnes étaient atteintes d'un cancer parce qu'elles avaient consommé du café. En fait, pour le savoir, il faut disposer d'études dont les données sont obtenues non pas en comparant des indicateurs populationnels, mais plutôt en observant des individus, comme le font les études de cohorte(s) ou cas-témoins.

14.1.1 Les particularités des études cas-témoins et de cohorte(s)

Les études cas-témoins

Dans les études cas-témoins, c'est la présence et l'absence de la maladie qui déterminent le choix des sujets et leur appartenance aux groupes d'étude. On sélectionne d'abord des personnes atteintes de la maladie étudiée, qui forment le groupe des cas, et on les

compare à des personnes qui n'en sont pas atteintes, lesquelles constituent le groupe des témoins. Les cas doivent être représentatifs de l'ensemble des cas dans la population cible. Les témoins doivent aussi être représentatifs de l'ensemble des autres personnes dans la population cible. Les témoins et les cas proviennent donc de la même population.

Par exemple, on pourra sélectionner les cas grâce à un registre provincial de cancer et procéder au recrutement de témoins par l'entremise d'une sélection aléatoire d'individus (en excluant les personnes atteintes du cancer étudié) de cette même province. L'étude s'appelle alors «étude cas-témoins populationnelle», que les Anglo-Saxons nomment «*population-based case-control study*». Souvent, il est difficile de respecter ce critère à la lettre. Ainsi, il peut s'avérer impossible de recruter un échantillon représentatif des nouveaux cas dans une population donnée. Dans une telle situation, l'échantillon pourra, par convenance, être constitué de personnes qui ont un diagnostic de la maladie étudiée en milieu hospitalier. On peut également éprouver des difficultés dans le choix des témoins lorsque la population d'où proviennent les cas est mal définie ou lorsqu'il s'avère impossible de tirer de cette population un échantillon aléatoire pour constituer le groupe des témoins. On tente alors de respecter au mieux le critère de recrutement des témoins dans la population d'origine en sélectionnant ceux-ci parmi les patients admis au même hôpital que les cas, mais pour une autre maladie. On parle alors d'une «étude cas-témoins en milieu hospitalier», que les Anglo-Saxons désignent par «*hospital-based case-control study*». Il existe aussi un type d'étude cas-témoins qu'on appelle imbriquée, ou «*nested case-control study*», qui a la particularité de puiser ses cas et ses témoins au sein d'une étude de cohorte existante. Une fois les cas et les témoins sélectionnés, on cherchera une ou plusieurs expositions qui pourraient expliquer l'apparition de la maladie.

Les études de cohorte(s)

Dans les études de cohorte(s), les sujets sont choisis en fonction de leur exposition à l'un des facteurs qu'on croit susceptibles d'induire la maladie. Les groupes à l'étude sont donc définis par leur exposition. En suivant l'évolution de l'état de santé des sujets faisant partie de chaque groupe d'étude, on peut surveiller la survenue de la maladie. Lorsque la période d'observation commence après le début de l'étude, on qualifie la cohorte de «prospective». Si, par contre, on cherche à cerner dans le passé d'une population un facteur de risque présumé, et que la période d'observation précède le début de l'étude, on qualifie la cohorte d'«historique» ou de rétrospective.

À retenir

- **Étude de cohorte(s)** : Les sujets sont recrutés sur la base de l'exposition. La fréquence de la maladie chez les sujets exposés est comparée à celle chez les sujets non exposés. L'étude peut être rétrospective (historique) ou prospective.
- **Étude cas-témoins** : Les sujets sont recrutés sur la base de la maladie. La fréquence de l'exposition (cote) chez les cas est comparée à celle chez les témoins. L'étude peut être à base populationnelle ou hospitalière, ou parfois imbriquée dans une cohorte existante.

14.1.2 L'exemple du café et du cancer du pancréas

Voyons comment, à partir d'un relevé de la littérature, Lisa peut s'appuyer sur les résultats des études d'observation pour bien expliquer à sa patiente Geneviève le lien entre la consommation de café et le cancer du pancréas. Elle a d'abord révisé les études cas-témoins, qui permettent d'étudier simultanément et assez rapidement plusieurs facteurs de risque pour une maladie donnée. Lisa s'assure de couvrir l'ensemble des études publiées sur le sujet pour bien saisir aussi l'évolution de la recherche sur la question.

L'une des premières études de ce genre a été menée par MacMahon⁷, à Boston, entre 1974 et 1979. Les chercheurs ont comparé 369 patients victimes du cancer du pancréas à 644 témoins hospitalisés, mais qui ne souffraient pas de maladies pancréatiques ou biliaires ni de maladies reconnues comme ayant un lien avec le tabac ou l'alcool. Les auteurs de l'étude en sont arrivés à la conclusion que le risque de cancer était plus élevé chez les hommes buveurs de café décaféiné, sans qu'il y ait toutefois de relation entre l'augmentation du risque et la quantité consommée. Le risque était également élevé chez les consommatrices de café, mais, cette fois, il y avait une nette corrélation entre la quantité consommée et l'augmentation du risque. Ces résultats demeuraient significatifs même après ajustement selon une variable confondante possible, la consommation de tabac. On notait en fait une faible association avec la consommation de tabac, mais aucun lien avec celle d'alcool. En revanche, d'autres études n'ont pas mis en évidence une telle association entre le cancer du pancréas et le café, mais plutôt un lien entre ce type de cancer et la consommation de tabac ou d'alcool. Dans la monographie de l'International Agency for Research on Cancer (IARC) publiée en 1991, des 21 études cas-témoins recensées sur la relation entre le cancer du pancréas et le café, 10 ont permis de trouver une association positive modérée⁸. Par contre, la plupart des études cas-témoins publiées après le rapport de l'IARC n'ont pas établi d'association entre le café et le cancer du pancréas⁹. Plusieurs problèmes liés aux études cas-témoins affaiblissaient les conclusions de celles qui avaient trouvé une association positive. Ainsi, dans plusieurs études, notamment dans celle de MacMahon, les chercheurs avaient exclu les témoins qui souffraient d'une maladie pouvant être attribuable au tabagisme. Or, un même individu est souvent fumeur et buveur de café. En excluant les fumeurs, on diminuait donc artificiellement la proportion des buveurs de café parmi les témoins créant ainsi un biais de sélection.

Dans la recherche des causes du cancer du pancréas, les difficultés tenaient surtout au fait que c'était souvent un proche (proxy) qui évaluait la consommation de café du cas décédé. Les proches ont pu surestimer la consommation de café des cas, produisant ainsi une surestimation de l'association entre le cancer et la consommation de café. De plus, les entrevues n'étaient pas effectuées en aveugle. Sachant qu'ils sont en présence d'un cas, les intervieweurs peuvent faire des efforts additionnels pour chercher l'information et mieux déclarer l'exposition chez ces derniers (biais d'information différentielle). Donc, pour obtenir des conclusions valides, on a dû effectuer des études de cohorte(s). Elles ont généralement mené à la conclusion qu'il n'y avait pas d'association entre la consommation de café et le cancer du pancréas. Le tabagisme était le facteur de risque le plus reconnu. Par exemple, une des récentes études de cohorte montre qu'il n'y a pas de lien entre le café et le cancer du pancréas¹⁰. Les auteurs de cette étude de cohorte ont examiné l'association entre la consommation

de café et l'incidence du cancer du pancréas parmi 457 366 adultes américains âgés de 50 à 71 ans chez qui l'on a mesuré la consommation de café au recrutement et au suivi pour un total de 4 155 256 personnes-année. Le nombre total de cas incidents de cancer du pancréas dans la cohorte a été de 1 541. Les auteurs n'ont pas trouvé de différence statistiquement significative entre la consommation de café et l'incidence du cancer du pancréas.

14.2 L'analyse critique des études cas-témoins et de cohorte(s)

L'utilisation d'une grille d'évaluation permet de faire une analyse critique rigoureuse d'un article scientifique, afin de juger de la qualité et de la validité d'une étude cas-témoins ou d'une étude de cohorte(s)¹¹. Le tableau 14.1 présente la traduction française de la liste de vérification élaborée par l'initiative internationale STROBE¹². Cette liste énumère les informations qui devraient se retrouver dans un article scientifique rapportant les résultats d'une étude d'observation telles les études à visée étiologiques cas-témoins et les études à visée étiologique de cohorte(s).

TABLEAU 14.1 ■ La liste de contrôle STROBE

	Item n°	Recommandation
Titre et résumé	1	(a) Indiquer dans le titre ou dans le résumé le type d'étude réalisée en termes couramment utilisés. (b) Fournir dans le résumé une information synthétique et objective sur ce qui a été fait et ce qui a été trouvé.
Introduction Contexte/justification	2	Expliquer le contexte scientifique et la légitimité de l'étude en question.
Objectifs	3	Citer les objectifs spécifiques, y compris toutes les hypothèses <i>a priori</i> .
Méthodes Conception de l'étude	4	Présenter les éléments clés de la conception de l'étude en tout début de document.
Contexte	5	Décrire le contexte, les lieux et les dates pertinentes, y compris les périodes de recrutement, d'exposition, de suivi et de recueil de données.
Population	6	(a) <i>Étude de cohorte</i> – Indiquer les critères d'éligibilité, et les sources et méthodes de sélection des sujets. Décrire les méthodes de suivi. <i>Étude cas-témoin</i> – Indiquer les critères d'éligibilité et les sources et méthodes pour identifier les cas et sélectionner les témoins. Justifier le choix des cas et des témoins. →

TABLEAU 14.1 ■ La liste de contrôle STROBE (suite)

	Item n°	Recommandation
		<p><i>Étude transversale</i> – Indiquer les critères d'éligibilité et les sources et méthodes de sélection des participants.</p> <p>(b) <i>Étude de cohorte</i> – Pour les études appariées, indiquer les critères d'appariement et le nombre de sujets exposés et non exposés.</p> <p><i>Étude cas-témoin</i> – Pour les études appariées, indiquer les critères d'appariement et le nombre de témoins par cas.</p>
Variables	7	Définir clairement tous les critères de résultats, les expositions, les facteurs de prédiction, les facteurs de confusion potentiels et les facteurs d'influence. Indiquer les critères diagnostiques, le cas échéant.
Sources de données/mesures	8*	Pour chaque variable d'intérêt, indiquer les sources de données et les détails des méthodes d'évaluation (mesures). Décrire la comparabilité des méthodes d'évaluation s'il y a plus d'un groupe.
Biais	9	Décrire toutes les mesures prises pour éviter les sources potentielles de biais.
Taille de l'étude	10	Expliquer comment a été déterminé le nombre de sujets à inclure.
Variables quantitatives	11	Expliquer comment les variables quantitatives ont été traitées dans les analyses. Le cas échéant, décrire quels regroupements ont été effectués et pourquoi.
Analyses statistiques	12	<p>(a) Décrire toutes les analyses statistiques, y compris celles utilisées pour contrôler les facteurs de confusion.</p> <p>(b) Décrire toutes les méthodes utilisées pour examiner les sous-groupes et les interactions.</p> <p>(c) Expliquer comment les données manquantes ont été traitées.</p> <p>(d) <i>Étude de cohorte</i> – Le cas échéant, expliquer comment les perdus de vue ont été traités.</p> <p><i>Étude cas-témoin</i> – Le cas échéant, expliquer comment l'appariement des cas et des témoins a été réalisé.</p> <p><i>Étude transversale</i> – Le cas échéant, décrire les méthodes d'analyse qui tiennent compte de la stratégie d'échantillonnage.</p> <p>(e) Décrire toutes les analyses de sensibilité.</p>

TABLEAU 14.1 ■ La liste de contrôle STROBE (suite)

	Item n°	Recommandation
Résultats Population	13*	(a) Rappporter le nombre d'individus à chaque étape de l'étude (par exemple : potentiellement éligibles, examinés pour l'éligibilité, confirmés éligibles, inclus dans l'étude, complètement suivis, et analysés). (b) Indiquer les raisons de non-participation à chaque étape. (c) Envisager l'utilisation d'un diagramme de flux.
Données descriptives	14*	(a) Indiquer les caractéristiques de la population étudiée (par exemple : démographiques, cliniques, sociales) et les informations sur les expositions et les facteurs de confusion potentiels. (b) Indiquer le nombre de sujets inclus avec des données manquantes pour chaque variable d'intérêt. (c) <i>Étude de cohorte</i> – Résumer la période de suivi (par exemple : nombre moyen et total).
Données obtenues	15*	<i>Étude de cohorte</i> – Rappporter le nombre d'évènements survenus ou les indicateurs mesurés au cours du temps. <i>Étude cas-témoin</i> – Reporter le nombre de sujets pour chaque catégorie d'exposition, ou les indicateurs du niveau d'exposition mesurés. <i>Étude transversale</i> – Reporter le nombre d'évènements survenus ou les indicateurs mesurés.
Principaux résultats	16	(a) Indiquer les estimations non ajustées et, le cas échéant, les estimations après ajustement sur les facteurs de confusion avec leur précision (par exemple : intervalle de confiance de 95 %). Expliciter quels facteurs de confusion ont été pris en compte et pourquoi ils ont été inclus. (b) Indiquer les valeurs bornes des intervalles lorsque les variables continues ont été catégorisées. (c) Selon les situations, traduire les estimations de risque relatif en risque absolu sur une période de temps (cliniquement) interprétable.
Autres analyses	17	Mentionner les autres analyses réalisées (par exemple : analyses de sous-groupes, recherche d'interactions, et analyses de sensibilité).
Discussion Résultats clés	18	Résumer les principaux résultats en se référant aux objectifs de l'étude. →

TABLEAU 14.1 ■ La liste de contrôle STROBE (suite)

	Item n°	Recommandation
Limitations	19	Discuter les limites de l'étude, en tenant compte des sources de biais potentiels ou d'imprécisions. Discuter du sens et de l'importance de tout biais potentiel.
Interprétation	20	Donner une interprétation générale prudente des résultats compte tenu des objectifs, des limites de l'étude, de la multiplicité des analyses, des résultats d'études similaires, et de tout autre élément pertinent.
« Généralisabilité »	21	Discuter la « généralisabilité » (validité externe) des résultats de l'étude.
Autre information Financement	22	Indiquer la source de financement et le rôle des financeurs pour l'étude rapportée, le cas échéant, pour l'étude originale sur laquelle s'appuie l'article présenté.

*Indiquer l'information séparément pour les cas et les témoins dans les études cas-témoins et, le cas échéant, pour les groupes exposés et non exposés dans les études de cohorte et les études transversales.

Remarque : Un article d'élaboration et d'explication traite chaque item de la liste de contrôle et indique le cadre méthodologique de référence accompagné d'exemples publiés dont la rédaction est claire. La liste de contrôle STROBE s'utilise mieux à l'aide de cet article (disponible gratuitement sur les sites Web de PLoS Medicine - <http://www.plosmedicine.org/>, Annals of Internal Medicine - <http://www.annals.org/>, et Epidemiology - <http://www.epidem.com/>).

Source : Cet article a déjà été publié sous la référence : « Gedda, M. Traduction française des lignes directrices STROBE pour l'écriture et la lecture des études observationnelles », *Kinesither Rev*, 2015, 15(157): 34-8. © 2014 Elsevier Masson SAS. Tous droits réservés. *Open access under CC BY-NC-ND license.*

14.2.1 L'analyse critique d'une étude cas-témoins

Voici un exemple d'analyse critique d'une étude cas-témoins sur les facteurs de risque du cancer du pancréas menée par Ghadirian, Simard et Baillargeon entre 1984 et 1988 au Québec. Cette analyse critique est effectuée en suivant l'ordre de présentation des énoncés dans la liste de vérification présentée dans le tableau 14.1 en regroupant parfois certains éléments pour éviter les répétitions.

Le titre et le résumé

Le résumé d'un article présente en quelques lignes les principaux éléments de l'étude. Il permet au lecteur d'exercer rapidement un premier jugement critique. Le résumé de l'étude de Ghadirian et de ses collègues se lit ainsi :

Une étude cas-témoins de population portant sur le cancer du pancréas a été réalisée dans le Grand Montréal entre 1984 et 1988. Un total de 179 cas et de 239 témoins choisis dans la population ont été soumis à une entrevue. Cette étude faisait partie du programme SEARCH de l'Agence internationale de recherche sur le cancer de l'Organisation mondiale de la Santé (International Agency for Research on Cancer of the World Health Organization). Une forte association a été observée entre la consommation totale de tabac et le risque de cancer du pancréas (rapport

de cotes [RC] de 3,76 ; intervalle de confiance [IC] à 95 % de 1,80 à 7,83). Le RC pour les fumeurs du dernier quintile était de 5,15 comparativement à 3,99 pour les ex-fumeurs. Ceux qui consommaient de l'alcool présentaient, en général, un moindre risque que les non-consommateurs d'alcool. Les consommateurs de café étaient à moindre risque que les non-consommateurs, particulièrement si le buveur consommait le café durant le repas plutôt que l'estomac vide.

Bien qu'il ne soit pas structuré, ce résumé passe néanmoins en revue les points importants de l'article : le type d'étude (devis) réalisée, ici une étude cas-témoins, ce qui a été fait, ses principaux résultats et les principales conclusions qui en ont été tirées.

L'introduction

On devrait trouver dans l'introduction d'un article scientifique un bref rappel de ce qu'on sait sur le sujet étudié, sur le contexte scientifique qui vient légitimer l'étude, de même qu'une description claire de ses objectifs et des hypothèses.

Le contexte/la justification Dans l'étude cas-témoins qui nous intéresse, les auteurs présentent brièvement ce qui ressort de la trentaine d'articles qu'ils ont consultés sur l'association entre la cigarette, l'alcool, le café et le cancer du pancréas. L'association entre le tabagisme et le cancer du pancréas semble faire l'unanimité. Le lien entre l'alcool et ce cancer n'a pas été confirmé dans toutes les études, encore moins celui entre le café et ce type de cancer.

Les objectifs Les auteurs ne précisent pas clairement leurs objectifs, en quoi leur étude diffère des autres études. On déduit qu'ils veulent vérifier à leur tour les liens non confirmés par les études antérieures entre l'alcool, le café et le cancer du pancréas. À ce chapitre, leur introduction présente des lacunes.

La méthode

La conception de l'étude Le type d'étude (le devis) est mentionné dans le résumé, mais il n'est ni repris ni abordé ailleurs dans l'article, ce qui constitue une lacune.

Le contexte Il est important de bien comprendre le contexte de l'étude pour juger de la validité interne et de la validité externe de celle-ci. L'information sur les lieux, sur les dates, dont les périodes de recrutement, d'exposition, de suivi et de collecte des données auprès de la population à l'étude, est présentée de la façon suivante. Cette information doit permettre de bien identifier la population cible. Ghadirian et ses collègues rapportent ce qui suit :

De 1984 à 1988, tous les patients âgés de 35 ans à 79 ans ayant reçu un diagnostic clinique ou histologique de cancer du pancréas, des voies ou de la vésicule biliaires ont été ciblés par les bureaux d'admission, les services de pathologie, de gastroentérologie et d'archives médicales de 19 hôpitaux francophones du Grand Montréal [...]. Les sujets témoins, assortis aux cas selon l'âge (selon des tranches d'âge de cinq ans), le sexe et le lieu de résidence, ont été sélectionnés aléatoirement à partir d'une liste de numéros de téléphone. L'indicatif régional et les trois premiers chiffres correspondant à l'indicatif local du numéro de téléphone de chacun des cas ont été appariés à un numéro de téléphone choisi par tirage au sort.

Concernant le suivi et le recueil de données, les auteurs expliquent que les participants à l'étude sont tous les francophones qui ont pu être interrogés avec l'accord du médecin traitant. L'entrevue a été réalisée avec le patient lui-même lorsque c'était possible. Dans le cas contraire, parce que la personne était trop malade ou était décédée, les chercheurs se sont adressés à des proches. Quatre personnes ont été exclues à la suite d'une erreur de diagnostic (diagnostic clinique infirmé à la pathologie finale). Au bout du compte, sur 447 cas identifiés, 179 ont participé à l'étude.

Quoique les auteurs ne parlent pas directement de la population cible ni de la population à l'étude, on peut croire que la population cible était l'ensemble de la population francophone du Québec âgée de 35 ans à 79 ans et que la population à l'étude était la population francophone du Grand Montréal. D'ailleurs, le sous-titre de leur article est « Une étude cas-témoins de population au Québec, Canada ». On aurait aimé avoir plus de précision sur les raisons du choix de la population à l'étude.

La population Il est important de bien définir les critères d'éligibilité, les sources et les méthodes de sélection des sujets à l'étude pour évaluer la présence de biais de sélection, ainsi que les critères d'appariement et le nombre de cas et de témoins pour vérifier la présence de biais de confusion. La description du mode de sélection des sujets (cas et témoins), des critères d'exclusion et d'inclusion et des non-réponses ou des refus aide à détecter la présence de biais de sélection. La comparabilité des groupes étudiés permet de détecter un risque de biais de confusion. S'il y a un déséquilibre dans les caractéristiques des cas et des témoins (par exemple, les cas proviennent d'un milieu plus défavorisé que celui des témoins), il sera possible que cette variable « non équilibrée » entre les groupes joue un rôle de confusion. C'est à l'étape de l'analyse que l'on jugera si les auteurs ont contrôlé adéquatement le potentiel de biais de confusion.

L'étude cas-témoins est particulièrement « vulnérable » aux biais de sélection, et la validité de ses résultats dépend en grande partie des méthodes utilisées pour sélectionner les cas et les témoins. Une méthode de sélection adéquate doit fournir aux chercheurs l'assurance que les cas et les témoins retenus forment un échantillon représentatif de la population à l'étude, c'est-à-dire que chaque individu doit avoir une chance égale – une probabilité égale – d'être sélectionné. Seule une méthode par tirage au sort, où toutes les personnes sélectionnées acceptent de participer, satisfait à cette exigence. Cette méthode permet d'éviter que la sélection des cas et des témoins soit influencée par l'exposition aux facteurs étudiés. Dans l'article de Ghadirian et ses collègues, on trouve à la section « Méthode » une description de la démarche utilisée pour sélectionner les cas et les témoins :

Entre juin 1984 et juin 1988, 447 cas de cancer du pancréas ont été identifiés dans 19 hôpitaux francophones de la région de Montréal. De ce nombre, 144 cas (32 %) ont été exclus de l'étude pour les raisons suivantes : dans 31 cas, la personne ne se situait pas dans les limites du groupe d'âge étudié (de 35 à 79 ans) ; dans 71 cas, la personne était décédée plus de un an avant la date prévue de l'entrevue ; et dans 42 cas, la personne résidait à l'extérieur de la région de Montréal. Des 303 cas admissibles, 179 personnes (59 %) ont été interviewées. Cent vingt-quatre personnes n'ont pu être interviewées pour les raisons suivantes : 32 médecins traitants n'ont pas accordé aux chercheurs l'autorisation d'interviewer

leurs patients ; 29 patients ou proches d'un patient ont refusé de participer à une entrevue ; 21 médecins traitants n'ont pas répondu à la demande faite par les chercheurs pour interviewer leurs patients ; 20 patients n'ont pu être contactés ; il n'y avait pas de répondant fiable pour 18 personnes décédées moins de un an avant l'entrevue ; enfin, 4 cas ont été exclus parce que le diagnostic de cancer du pancréas était erroné.

On y trouve également une description de la méthode de sélection des sujets témoins :

Les sujets témoins, assortis aux cas selon l'âge (selon des tranches d'âge de cinq ans), le sexe et le lieu de résidence, ont été sélectionnés aléatoirement à partir d'une liste de numéros de téléphone. L'indicatif régional et les trois premiers chiffres correspondant à l'indicatif local du numéro de téléphone de chacun des cas ont été appariés à un numéro de téléphone choisi par tirage au sort. Le numéro de téléphone était composé ; lorsqu'un témoin admissible acceptait d'être interviewé, on convenait d'un rendez-vous pour une entrevue [...]. Bien que cette méthode se soit avérée adéquate pour l'étude pilote, il s'est révélé très difficile de sélectionner ainsi l'ensemble des témoins, car on s'est heurté à un très fort taux de refus de participer. Nous avons été obligés de modifier la méthode de sélection des témoins : les témoins ont donc été sélectionnés à partir de l'annuaire téléphonique dans lequel le cas correspondant était inscrit (à noter que tous les cas étudiés étaient inscrits dans l'annuaire téléphonique, seulement 1 % des familles de la région de Montréal n'ayant pas le téléphone). Une page de l'annuaire téléphonique a été choisie au hasard, et le nom et l'adresse de 10 personnes ayant le même indicatif local (soit les trois premiers numéros) que celui du cas ont été retenus. Une lettre dans laquelle on expliquait le but de l'étude a par la suite été acheminée à ces adresses. Environ une semaine plus tard, on a téléphoné à ces numéros pour vérifier s'il s'y trouvait un témoin dont l'âge et le sexe correspondaient à ceux du cas à appairer. Lorsque c'était le cas et que la personne jointe acceptait de participer à l'étude, une entrevue était réalisée à son domicile. Lorsque la personne refusait d'être interviewée, on passait au numéro suivant. S'il y avait plus d'une personne admissible à un numéro donné, cette information était conservée dans la banque de données pour un usage ultérieur.

Pour être admissible, un témoin devait avoir été vivant au moment où le diagnostic de cancer avait été posé chez le patient auquel on cherchait à appairer ce témoin. Les témoins vivants ont été interviewés directement. Lorsque le témoin était décédé entre le diagnostic de cancer du cas assorti et la date de l'entrevue, un proche du témoin était interviewé. Au total, 239 témoins ont été interviewés à leur domicile.

Bien que la méthode de sélection des cas et des témoins semble adéquate, il existe certaines limites quant à la sélection finale. Tous les nouveaux cas de cancer du pancréas survenus au cours d'une période de 4 ans dans les 19 hôpitaux francophones de la région de Montréal sélectionnés ont probablement été identifiés. Les sujets témoins ont été choisis aléatoirement dans la population francophone résidant dans la région de Montréal. Les cas et les témoins ont dû satisfaire aux mêmes critères d'inclusion et d'exclusion : les personnes âgées de 35 à 79 ans ont été incluses dans l'étude ; les patients ne résidant pas dans la région de Montréal et ceux décédés plus de un an avant la date de l'entrevue ont été exclus.

Par contre, la population d'où proviennent les cas n'est pas bien définie, puisque 21 hôpitaux de soins de courte durée et plusieurs centres de soins prolongés n'ont pas été inclus dans l'étude. Ces centres reçoivent une bonne proportion de patients francophones. Il serait étonnant que tous les francophones atteints du cancer du pancréas aient été admis dans les hôpitaux choisis. Il est aussi possible que le niveau de la sur-spécialisation des différents hôpitaux ait influencé le spectre de sévérité des cas inclus.

Les variables et les sources de données/mesures Les auteurs indiquent que les cas ont été identifiés à l'aide du diagnostic histologique. Toutes les informations auprès des sujets à l'étude ont été obtenues à l'aide d'un questionnaire standardisé selon un protocole d'entrevue normalisé réalisé par des interviewers formés. Ce questionnaire a été conçu, évalué et testé sous la supervision de l'IARC. Les auteurs indiquent les objets couverts par les 40 questions du questionnaire. Ils mentionnent aussi avoir utilisé un deuxième questionnaire portant sur les facteurs de risque nutritionnels sans toutefois en mentionner la référence.

Les biais Les sources potentielles de biais doivent être analysées avec minutie afin d'évaluer la validité de l'étude.

- Non-réponse et refus de participer

Parmi les 303 cas éligibles à l'étude de Ghadirian et ses collègues, 60 % ont pu être interviewés. Pour 11 % des cas, l'autorisation du médecin n'a pas été obtenue, 9 % ont refusé de participer à l'étude, 7 % n'ont pu être joints, 6 % sont décédés et pour 7 % des cas, le médecin n'a pas répondu à la demande de permission d'interviewer le patient. Un nombre non spécifié, mais probablement encore plus important de témoins ont également refusé d'être interviewés. Les auteurs ont dû changer de méthode de recrutement des témoins en cours d'étude à cause d'un taux de refus très élevé. Une telle situation peut introduire un biais de sélection et invalider les résultats. Il est possible que ceux qui ont refusé de participer à l'étude n'aient pas été exposés aux facteurs de la même façon que les sujets qui ont accepté d'y participer. Ainsi, ces derniers ont pu être au courant de rumeurs sur le lien possible entre le café et le cancer du pancréas : on peut envisager que les gros buveurs de café – cas admissibles fortement exposés – ont systématiquement refusé de participer à l'étude. Dans cette situation, la cote de l'exposition parmi les cas participants sera sous-estimée. De la même façon, il est possible que les témoins admissibles les moins exposés, c'est-à-dire ceux qui buvaient peu de café ou pas du tout, aient refusé de participer à l'étude. La cote de l'exposition des témoins participants sera alors surestimée. En conséquence, la mesure du RC pourrait être surestimée ou sous-estimée. En somme, même une méthode adéquate de sélection des cas et des témoins peut être invalidée par une forte proportion de refus de participer.

Il faut cependant ajouter qu'une proportion différente de refus de participer chez les cas et chez les témoins n'entraîne pas automatiquement un biais de sélection. En effet, il se peut que la proportion des sujets exposés et des sujets non exposés parmi les gens qui ont refusé de participer soit la même que celle parmi ceux qui ont accepté de participer. Dans cette situation, il n'y aurait pas de biais de sélection. Pour le démontrer, imaginons une étude cas-témoins portant sur le lien entre le cancer du

poumon et le tabagisme où 200 cas et 400 témoins ont été identifiés à l'aide d'une procédure aléatoire. Tous ont accepté de participer. Les observations pourraient être résumées dans un tableau de contingence qui ressemblerait à celui-ci :

	Tabagisme +	Tabagisme -	Total
Cas	120	80	200
Témoins	80	320	400
Total	200	400	600

RC = 6

La mesure d'association obtenue, soit un RC de 6 $((120/80)/(80/320))$, n'est pas biaisée puisque tous les cas et tous les témoins sélectionnés participent à l'étude. La proportion de cas fumeurs est de 60 % (120/200), et la proportion de témoins fumeurs est de 20 % (80/400). Maintenant, supposons que 30 % des cas (60) et 40 % des témoins (160) refusent de participer à l'étude. Manifestement, la proportion des cas et des témoins qui ont refusé de participer est alors différente. Supposons encore que 60 % des cas (36) et 20 % des témoins (32) qui refusent de participer sont des fumeurs (les mêmes proportions de fumeurs que celles dans la réalité). Les observations relatives aux cas et aux témoins qui ont accepté de participer à l'étude pourraient être cette fois résumées dans un tableau de contingence comme celui-ci :

	Tabagisme +	Tabagisme -	Total
Cas	84 (120 - 36)	56 (80 - 24)	140
Témoins	48 (80 - 32)	192 (320 - 128)	240
Total	132	248	380

RC = 6

On voit bien que les refus de participer lors de la sélection des cas et des témoins n'ont pas changé le RC ; le RC mesuré est le même à cause d'une répartition égale de l'exposition chez les participants et les non-participants à l'étude. Donc, le RC n'est pas biaisé. Il l'aurait cependant été si la répartition de l'exposition parmi les personnes refusant de participer n'avaient pas été égale à celle observée parmi les participants.

■ La prévention des biais liés aux refus de participer

Certes, lorsqu'on mène une étude, on ne connaît pas l'exposition des personnes invitées à y participer mais qui refusent de le faire. Par conséquent, comment les chercheurs doivent-ils agir ? Y a-t-il aussi un moyen de prévenir ou de contrôler le biais possible associé aux refus de participer ? Dans l'article de Ghadirian et de ses collègues, on ne trouve ni information ni discussion sur les efforts consentis pour prévenir ce biais de sélection. Or, la prévention du biais de sélection au moment du choix des participants demeure la conduite à privilégier, puisqu'il est impossible de le contrôler à l'étape de l'analyse. Cependant, la prévention n'est possible que dans

une certaine mesure. Ainsi, des stratégies peuvent être mises au point afin d'inciter les médecins traitants à donner leur accord pour que leurs patients participent à l'étude et d'inciter ces derniers à s'y engager. Par ailleurs, déterminer si un patient décédé est un participant adéquat s'avère une tâche difficile, parfois irréalisable.

Si les mesures préventives n'ont pas permis d'éviter les refus de participer à l'étude, on peut toujours essayer d'estimer l'impact potentiel du biais de sélection à l'étape de l'analyse avec des analyses dites de sensibilité. Pour ce faire, il faut cependant connaître les caractéristiques d'exposition des cas et des témoins qui ont refusé de participer. Ainsi, les cas qui ont refusé de participer sont-ils plus nombreux à fumer ou à boire du café que ceux qui ont accepté? Il est souvent difficile d'estimer l'exposition à des facteurs de risque chez les personnes qui ont refusé de participer à l'étude. On doit émettre des hypothèses. Par exemple, on peut refaire les analyses en prenant en considération les refus et en leur attribuant les pires caractéristiques d'exposition. On peut, par exemple, considérer que les cas qui ont refusé de participer sont tous des non-fumeurs et que les témoins, par contre, sont tous de gros fumeurs. À partir de cette hypothèse, on refait le calcul du RC. Il est clair qu'il va alors se rapprocher de la valeur nulle ($RC = 1$). Mais si l'intervalle de confiance ne contient pas la valeur nulle, on aura prouvé que ce biais possible ne modifie pas les conclusions de l'étude, même dans les pires scénarios.

Malheureusement, on ne peut écarter la possibilité que l'étude de Ghadirian et de ses collègues présente un sérieux biais de sélection du seul fait qu'on a noté un nombre très élevé de refus de participer et qu'aucune information n'est présentée pour évaluer l'impact de ces refus.

La faible participation des témoins contactés nous amène à discuter d'autres stratégies qui auraient pu être utilisées pour sélectionner les témoins. Par exemple, on aurait pu choisir les témoins parmi les patients du même âge et du même sexe, admis au même hôpital, dans la même semaine que les cas pour un autre diagnostic que celui de cancer du pancréas. Comme mentionné plus haut, les études cas-témoins peuvent se classer en deux groupes selon le procédé utilisé pour identifier et choisir les cas et les témoins: l'étude cas-témoins de population et l'étude cas-témoins en milieu hospitalier. Certains auteurs considèrent que l'étude cas-témoins de population est préférable à l'étude cas-témoins en milieu hospitalier parce qu'elle fournit l'assurance que les cas et les témoins proviennent d'une même population. Par contre, lorsque le refus de participer des témoins sélectionnés dans la population remet en question la validité interne même de l'étude, il peut s'avérer préférable de procéder à une étude en milieu hospitalier, qui présente alors des avantages certains: les sujets sont plus faciles à joindre et souvent plus coopératifs, et l'accès au dossier médical facilite la collecte de l'information sur la maladie et les examens diagnostiques. Le principal inconvénient de ce type d'étude vient du fait qu'il est alors plus difficile de bien cerner la population à l'étude, même si, en principe, les témoins hospitalisés proviennent de la même population que les cas hospitalisés, à moins que l'hôpital ne soit un centre de référence pour prendre en charge la maladie à l'étude (celle des cas).

Enfin, bien que l'étude cas-témoins en milieu hospitalier présente certains avantages d'ordre pratique, il demeure difficile de s'assurer que les témoins sélectionnés sont

représentatifs de l'ensemble des témoins (non-malades) dans la population. Pour cette raison, il faut lui préférer, quand c'est possible, l'étude cas-témoins de population. Le tableau 14.2 passe en revue les avantages respectifs des études cas-témoins de population et des études cas-témoins en milieu hospitalier.

TABEAU 14.2 ■ Les avantages respectifs des études cas-témoins de population et des études cas-témoins en milieu hospitalier

Caractéristiques	Études cas-témoins de population	Études cas-témoins en milieu hospitalier
Population à l'étude	Bien définie	Souvent mal définie
Provenance des cas et des témoins	Facilement contrôlable	Difficilement contrôlable
Représentativité des témoins	Représentation adéquate des non-malades dans la population à l'étude	Représentation incertaine des non-malades dans la population à l'étude
Accessibilité et coopération des sujets de l'étude	Cas et témoins souvent difficiles à recruter	Cas et témoins sur place plus enclins à participer à l'étude
Collecte de l'information	Peut être ardue	Facilitée par l'accessibilité des sujets et de leurs dossiers médicaux

■ Les décès survenus avant ou durant l'étude

Il pourrait aussi exister un biais d'échantillonnage (un biais de sélection particulier aux études cas-témoins). Le biais d'échantillonnage, ou biais de survie sélective, fait référence à une situation dans laquelle les personnes asymptomatiques et celles ayant connu des épisodes de maladies courtes et fatales sont moins susceptibles d'être admises dans l'étude. Il peut en résulter une erreur dans la mesure d'association, et, selon les circonstances, la vraie valeur pourrait être surestimée ou sous-estimée. Le biais d'échantillonnage peut être évité en s'assurant de n'admettre dans l'étude que les nouveaux cas, les cas incidents, à l'exclusion des cas connus, les cas prévalents.

Dans la section «Méthode» de l'étude de Ghadirian et de ses collègues, on peut lire ceci :

Entre juin 1984 et juin 1988, 447 cas de cancer du pancréas ont été dépistés dans 19 hôpitaux francophones de la région de Montréal.

Cette description laisse soupçonner que l'étude porte sur de nouveaux cas de cancer (l'ensemble des cas incidents), mais on ne peut porter un jugement définitif, puisque les auteurs ne précisent pas si les personnes déjà hospitalisées pour un cancer au début de l'étude, soit les cas prévalents, ont été incluses ou non dans celle-ci. Si elles ont été incluses dans l'étude, il pourrait s'agir de personnes moins exposées à des facteurs de risque comme la consommation de tabac et de café et pour qui le pronostic pourrait être meilleur, ce qui entraînerait une sous-estimation de la mesure d'association. Par ailleurs, 16 % des cas potentiels étaient décédés plus de un an

avant l'entrevue. Cela sème le doute sur l'inclusion exclusive de cas incidents. Il est possible que les cas présentant le pire pronostic n'aient pas été inclus dans l'étude, ce qui produirait un biais de sélection.

- La surveillance médicale

Un autre type de biais de sélection susceptible de se présenter dans les études cas-témoins est le biais de surveillance. Ce type de biais peut survenir lorsque la maladie étudiée est asymptomatique ou peu symptomatique. Les cas ont alors plus de chances d'être détectés chez les personnes qui font déjà l'objet d'une surveillance médicale attentive. En voici un exemple souvent cité dans la littérature scientifique.

EXEMPLE 14.1

Il y a plusieurs années, l'augmentation du risque de cancer de l'endomètre chez les femmes recourant aux œstrogènes seuls pour contrer les symptômes de la ménopause a fait l'objet de publications. Comme ces hormones entraînent des saignements utérins et que, pour cette raison, les femmes qui en prennent font l'objet d'une surveillance médicale particulière, on a émis l'hypothèse voulant que la surveillance médicale accrue ait permis de découvrir des cancers asymptomatiques chez des femmes atteintes avant même qu'elles ne commencent à prendre des œstrogènes. De ce fait, on pouvait avoir surestimé la force de l'association entre la prise d'œstrogènes et le cancer de l'endomètre. Il a toutefois été établi depuis que le biais de surveillance n'expliquait pas l'association observée et que les œstrogènes constituaient effectivement un facteur de risque, raison pour laquelle ils ne sont plus prescrits seuls chez les femmes avec un utérus.

Dans l'étude qui nous intéresse, il risque peu d'y avoir un biais de surveillance. En effet, on s'intéresse au cancer du pancréas, un cancer dont les symptômes apparaissent rapidement, sont sévères et entraînent presque invariablement une consultation médicale et pour lequel la survie moyenne est très courte.

- La mesure de l'exposition et de la maladie

La mesure de l'exposition et de la maladie est aussi une source potentielle de biais. Il est important de comprendre la mesure de l'exposition et de l'issue pour juger de la présence d'un biais d'information. Dans la section « Méthode », les auteurs devraient donner les définitions de ce que constituent un cas et un témoin, un sujet exposé ou non exposé. Les instruments de mesure et leurs caractéristiques devraient également être décrits, de même que leurs modalités d'utilisation. Ces précisions permettent d'évaluer la présence de biais d'information.

Le biais d'information résulte d'une mesure inadéquate ou d'une observation incorrecte d'un phénomène chez les participants. Ce biais est différentiel lorsque l'erreur commise s'avère différente selon le groupe auquel appartiennent les sujets, comme les personnes malades et non malades ou les personnes exposées et non exposées. À l'inverse, ce biais est non différentiel lorsque l'erreur demeure la même dans tous les groupes de sujets ayant participé à l'étude.

Ainsi, quand on fait la lecture critique d'une étude, il faut se demander si les auteurs ont cherché à prévenir les biais d'information lors de la planification et de l'exécution de l'étude, ou bien s'ils les ont identifiés et s'ils ont vérifié leurs impacts au moment de l'analyse. Pour prévenir les biais différentiels, il faudra utiliser les mêmes instruments de mesure, administrés de la même façon, pour tous les groupes. Si les instruments sont administrés par des intervieweurs, ceux-ci ne devraient pas connaître le groupe d'étude des participants qu'ils rencontrent. Pour prévenir les biais non différentiels, on doit s'assurer que ces instruments sont de bonne qualité et bien calibrés. Enfin, les personnes qui les manipulent doivent avoir reçu une formation adéquate et savoir interpréter les résultats correctement. On trouve habituellement cette information dans la section « Méthode » d'un article.

■ La source d'information pour mesurer la maladie

La source d'information utilisée pour mesurer la maladie dans l'étude de Ghadirian et de ses collègues est susceptible de provoquer un biais d'information non différentiel parce que les cas n'ont pas tous été sélectionnés en fonction du même critère. Pour la plupart d'entre eux, le diagnostic de cancer a été confirmé par un examen histologique de la tumeur. Toutefois, 17 % des cas ont été identifiés en fonction d'indices cliniques ou radiologiques. En l'absence de confirmation histologique, on peut penser qu'un certain nombre de sujets non atteints du cancer du pancréas ont pu être inclus dans le groupe des cas (quoique aléatoire selon le statut d'exposition), créant ainsi un biais d'information non différentiel.

■ L'appel à la mémoire : le biais de mémoire

Dans l'étude cas-témoins de Ghadirian et de ses collègues, le seul instrument de mesure utilisé pour estimer les expositions est un questionnaire normalisé. On affirme qu'il a été validé par l'IARC, qui a commandé l'étude. Les interviewers ont été bien formés et pouvaient compter sur un manuel pour les guider.

Les études cas-témoins présentent souvent un biais non différentiel appelé « biais de mémoire ». Plus lointaine est la période à laquelle remonte l'exposition, plus forte est la probabilité de la présence d'un biais de mémoire. Dans l'étude qui nous intéresse, ce biais a vraisemblablement joué un rôle, car on a interrogé les participants au sujet d'habitudes de consommation couvrant une vie entière. De plus, une forte proportion des personnes interrogées étaient des proches des cas ou des témoins. Leurs souvenirs devaient être d'autant plus imprécis.

Le biais d'information non différentiel a généralement pour effet de diminuer la force de l'association. Si les auteurs d'une étude concluent qu'il n'existe pas d'association entre un facteur et une maladie (c'est le cas ici en ce qui concerne la consommation de café), on peut penser que le biais non différentiel a peut-être masqué une association présente. Par contre, si l'on conclut dans la même étude à une association malgré la présence d'un biais non différentiel (c'est le cas ici pour la consommation de tabac et d'alcool), on doit se dire non seulement que l'association existe, mais qu'elle est probablement plus forte que celle qu'on a mise au jour. En somme, on peut donc affirmer que le fait de soupçonner la présence d'un biais d'information non différentiel dans une étude a pour effet de renforcer la crédibilité des conclusions positives et de mettre en doute les conclusions négatives.

La situation est différente dans le cas des biais différentiels, car ceux-ci peuvent avoir pour effet de diminuer ou d'augmenter la force de l'association. D'abord, pour prévenir ces biais, il faut utiliser les mêmes instruments de mesure pour les cas et pour les témoins, ainsi que pour les personnes exposées et pour les personnes non exposées, puis les administrer de la même manière pour les deux groupes. On peut aussi travailler en aveugle pour la collecte de l'information et pour l'analyse. On s'assure ainsi à tout le moins que s'il existe un biais d'information, il sera non différentiel.

Un point faible des études cas-témoins réside dans le fait que l'on doit se fier à la mémoire des gens. Or, la mémoire n'introduit pas qu'un biais non différentiel comme celui qu'on vient de voir. Elle est également à l'origine d'un autre biais, différentiel celui-là, appelé «biais de rappel». Il y a biais de rappel lorsque la personne malade, consciemment ou inconsciemment, fouille mieux dans sa mémoire pour mettre au jour des facteurs qui expliqueraient son état de santé. Évidemment, la personne en bonne santé n'a pas les mêmes préoccupations que celle qui est malade; ses souvenirs pourront donc être altérés. Dans l'étude de Ghadirian et de ses collègues, on n'aborde pas cette question, sans doute parce que les auteurs sont davantage préoccupés par une forme particulière de biais de rappel, celui dit «de l'informateur de seconde main». En effet, une faible proportion de malades (25 %) ont répondu eux-mêmes au questionnaire. En ce qui concerne les autres cas, on a obtenu l'information sur l'exposition en interrogeant un proche. Par contre, 83 % des témoins ont répondu directement au questionnaire. Il est certain que, pour des questions concernant la consommation d'alcool, de café ou de tabac, la réponse du proche risque fort de différer de celle qu'aurait donnée la personne en cause. Les auteurs en sont d'ailleurs bien conscients et en discutent longuement. En réalisant une analyse stratifiée selon le statut des répondants, c'est-à-dire selon que les cas ont répondu eux-mêmes au questionnaire ou que l'information sur l'exposition a été obtenue en interrogeant un proche, ils ont trouvé des résultats différents. Lorsqu'on ne considère que les réponses fournies par les cas, on obtient un RC de 3,04. Par contre, lorsqu'on prend en compte l'information donnée par les proches, le RC augmente à 5,35. Un biais de rappel risque donc fort d'avoir été introduit dans l'étude puisque les proches surestiment l'exposition au facteur de risque chez les cas.

■ L'attention portée aux cas et aux témoins

Une autre lacune des études cas-témoins tient au fait que les interviewers peuvent être influencés s'ils connaissent le groupe d'étude du participant. En effet, comme ils savent si la personne interviewée est un cas ou un témoin, ils peuvent être tentés d'interroger plus à fond les cas sur la question de l'exposition présumée. Ici, le questionnaire adéquatement normalisé a peut-être réduit ce biais au minimum. En général, c'est d'ailleurs le seul moyen de le prévenir, parce qu'il est difficile de mener les entrevues en aveugle, sauf si la maladie est peu symptomatique ou si toutes les entrevues se font avec des proches plutôt qu'avec les malades eux-mêmes, ou encore si les témoins sont sélectionnés parmi des personnes hospitalisées (donc malades) pour une autre raison.

La taille de l'étude On devrait expliquer ici comment a été déterminé le nombre de sujets à inclure dans l'étude. Dans l'étude de Ghadirian et de ses collègues, on ne mentionne pas toutes les raisons qui ont conduit à la taille des échantillons utilisés.

Bien sûr, 179 patients atteints d'un cancer du pancréas ont été utilisés parce que c'est ce qui restait des 447 cas identifiés au cours des 4 ans qu'a duré l'étude. Cependant, les auteurs ne discutent pas de l'influence de ce nombre sur la précision de l'étude. On n'y explique pas non plus pourquoi on a choisi 239 témoins. Par contre, théoriquement, on sait que le nombre de témoins dépassera rarement quatre pour un cas¹³. En effet, il s'agit là de la meilleure combinaison pour obtenir la plus grande précision possible au moindre coût. Au-delà de cette proportion, le nombre de témoins n'influence plus beaucoup la précision de l'étude. Par contre, l'étude a une puissance suffisante concernant l'association entre le tabac et le cancer du pancréas, puisque les risques relatifs sont statistiquement significatifs.

Les variables quantitatives Dans la section «Méthode», les auteurs ne présentent pas vraiment la façon dont les variables quantitatives ont été traitées. On trouve certaines informations sommaires dans la description des analyses statistiques réalisées. Les regroupements effectués pour analyser les variables quantitatives se trouvent essentiellement dans la présentation des résultats.

Les analyses statistiques (le plan d'analyse) En plus de décrire les analyses statistiques, les méthodes d'analyse par sous-groupes et les interactions, les auteurs doivent expliquer dans une étude cas-témoins comment l'appariement des cas et des témoins a été réalisé. En effet, l'un des éléments les plus importants lorsqu'on considère l'analyse statistique consiste à évaluer si le contrôle des facteurs de confusion potentiels a été pris en compte.

On sait que le biais de confusion est une erreur systématique induite par la présence d'un facteur associé tant à la maladie qu'au facteur d'exposition à l'étude. Il faut également que ce facteur n'appartienne pas à la chaîne causale de la maladie.

Dans la pratique, la comparaison des caractéristiques des cas et des témoins permet de suspecter la présence de variables de confusion lorsque ces variables sont distribuées inégalement entre les groupes. Dans l'étude de Ghadirian et de ses collègues, différents facteurs de confusion peuvent invalider les conclusions. Cette étude cas-témoins visait à établir un lien entre certaines habitudes de vie, comme le fait de fumer et de consommer du café et de l'alcool, et la survenue du cancer du pancréas. On peut penser à plusieurs variables qui sont associées autant à la maladie étudiée qu'à l'exposition (café, tabac, alcool) et qui pourraient venir biaiser la mesure de l'effet dans cette étude, comme l'âge.

On sait que l'âge peut être associé, de façon indépendante, à la fois à des habitudes de vie et au cancer du pancréas ; en effet, la fréquence de la plupart des cancers augmente avec l'âge. Le sexe est un autre exemple de facteur de confusion possible. Au moment de cette étude, le cancer du pancréas était plus fréquent chez les hommes que chez les femmes. Le sexe était donc un facteur associé à la maladie, mais aussi lié aux habitudes de vie. Ainsi, dans la plupart des sociétés, les femmes boivent moins d'alcool que les hommes. Donc, le sexe pouvait aussi être associé au facteur d'exposition. En conséquence, l'âge et le sexe sont des facteurs qui, s'ils sont distribués inégalement entre les cas et les témoins, peuvent introduire une distorsion dans la mesure d'association entre les habitudes de vie et le cancer du pancréas.

Le niveau socioéconomique est également souvent associé aux habitudes de vie. Ainsi, il est bien connu que les personnes de niveau socioéconomique peu élevé fument plus. Un niveau socioéconomique peu élevé est aussi un facteur associé à une plus grande fréquence de la plupart des cancers. Le groupe ethnique d'appartenance peut aussi constituer un facteur de confusion. Selon l'appartenance ethnique, les habitudes de vie peuvent varier. La consommation de tabac, d'alcool et de café diffère probablement chez les francophones, les anglophones et les allophones d'origine asiatique vivant au Québec, par exemple. De ce fait, et aussi peut-être pour des raisons génétiques, il est possible que la fréquence du cancer du pancréas varie en fonction du groupe ethnique d'appartenance. Si ces facteurs, à savoir le niveau socioéconomique et l'appartenance ethnique, ne sont pas distribués également entre les cas et les témoins, la mesure d'association pourra être biaisée. Nous conviendrons donc que l'âge, le sexe, le niveau socioéconomique et l'appartenance ethnique sont des facteurs de confusion potentiels qui pouvaient être présents dans l'étude.

Il existe plusieurs moyens de contrôler ou de réduire les biais de confusion. Ces moyens sont utilisés, selon le cas, au moment de l'échantillonnage (*a priori*) ou de l'analyse (*a posteriori*). Au cours de l'échantillonnage, on peut procéder par restriction : on exclut alors de l'étude les personnes qui sont exposées au facteur de confusion possible, ou encore on n'inclut que les individus d'une même catégorie de la variable de confusion potentielle (par exemple, uniquement des hommes ou des jeunes).

Pour contrôler ou réduire les biais de confusion *a priori*, on peut aussi procéder par assortiment (appariement). L'assortiment vise à rendre les groupes étudiés plus homogènes en ce qui concerne les facteurs de confusion possibles¹⁴. On associe alors à chacun des cas un ou des témoins qui s'apparentent à lui en ce qui a trait à certaines variables confondantes possibles. L'assortiment peut aussi se faire en fonction de certaines catégories de cas. Ghadirian et ses collègues ont utilisé cette méthode : chacun des cas a été assorti à un ou plusieurs témoins appartenant au même groupe d'âge, au même sexe et à la même zone résidentielle¹⁵. Lorsque l'assortiment se fait selon le rapport «un cas pour un témoin», on parle d'appariement. Pour une plus grande efficacité, l'assortiment exige qu'on utilise des méthodes statistiques particulières lors de l'analyse des données. Ainsi, l'assortiment peut augmenter la précision des mesures d'association et permet donc d'utiliser des échantillons de plus petite taille. Par contre, l'assortiment demande souvent beaucoup de temps et peut conduire à une perte d'information si des sujets doivent être écartés de l'analyse parce qu'ils n'ont pas été appariés.

Des méthodes de contrôle des biais de confusion peuvent par ailleurs être utilisées au moment de l'analyse, notamment la stratification et la modélisation statistique. Par exemple, les auteurs de l'étude qui nous intéresse ont utilisé une méthode d'analyse multivariée (*voir le chapitre 11*). Cette méthode est appropriée lorsqu'on veut contrôler simultanément plusieurs variables de confusion potentielles.

EXEMPLE 14.2

La description de l'analyse statistique présentée par Ghadirian et ses collègues et insérée dans la section des résultats est extrêmement sommaire. Les auteurs mentionnent qu'ils ont utilisé un modèle de régression logistique conditionnel et que les variables de confusion prises en compte sont l'âge, le sexe, le niveau de scolarisation et le type de répondant. Si l'on considère que les auteurs ont utilisé l'assortiment et la modélisation en prenant en compte la majorité des variables de confusion potentielles, on peut dire que les biais de confusion possibles ont été bien contrôlés dans cette étude. Cependant, les auteurs n'indiquent pas comment les données manquantes ont été traitées.

Les résultats

La population Comme nous l'avons lu précédemment à la section «Méthode» (étape 6, population), les auteurs rapportent le nombre d'individus à chaque étape de sélection et d'éligibilité. Ils indiquent les raisons de non-participation à chaque étape. Un diagramme de flux n'est cependant pas présenté.

Les données descriptives Les caractéristiques démographiques, outre le sexe et l'âge moyen des groupes, et les caractéristiques sociales et cliniques des sujets à l'étude ne sont pas décrites. Il s'agit d'une lacune importante. Un tableau détaillant ces informations aurait permis, par exemple, de comprendre les différences existantes entre les groupes.

Les données obtenues Par contre, l'étude contient des tableaux qui présentent une information détaillée sur le nombre de sujets pour chaque catégorie d'exposition tant pour l'exposition au tabac, à l'alcool et au café.

Les principaux résultats et les autres analyses Les auteurs présentent, pour les trois facteurs d'exposition étudiés et pour différents niveaux d'exposition, les nombres bruts de cas et de témoins, les rapports de cote ajustés et leur intervalle de confiance ainsi que la valeur p d'un test du khi carré permettant de juger si la différence entre les cas et les témoins est significative sur le plan statistique. Les rapports de cotes bruts ne sont pas présentés, mais les nombres bruts de cas et de témoins permettent de les calculer.

L'un des éléments importants à évaluer dans la section des résultats réside aussi dans la précision de l'étude, principalement déterminée par l'étendue des intervalles de confiance proposés. Plus un intervalle de confiance est large, moins la précision est grande. Cependant, il n'y a pas de normes absolues qui permettent de juger de cette précision. Dans l'étude de Ghadirian et de ses collègues, les intervalles de confiance rapportés sont les suivants :

- gros fumeur comparativement à non-fumeur: RC de 3,76 et ic de 1,8 à 7,83;
- gros buveur d'alcool comparativement à non-buveur d'alcool: RC de 0,65 et ic de 0,3 à 1,44;
- gros buveur de café comparativement à non-buveur de café: RC de 0,55 et ic de 0,19 à 1,62.

Par ailleurs, les auteurs présentent les résultats ajustés pour la confusion, mais ils ne tiennent pas compte de l'effet de modifications possibles dans leur analyse.

La discussion et la conclusion

Généralement, les auteurs discutent de leurs résultats et présentent leurs conclusions dans la dernière section de leur article. Cette section devrait présenter un résumé des principales découvertes et les mettre en lien avec la littérature existante. On devrait également y trouver une réflexion sur la validité interne, la validité externe et la signification clinique potentielle des résultats. Ces éléments correspondent aux points 18 à 21 de la grille d'analyse STROBE.

EXEMPLE 14.3

Dans l'étude qui nous intéresse, les auteurs discutent malheureusement très brièvement de leurs résultats. Voici quelques extraits de cette discussion.

Les résultats de cette étude en ce qui concerne l'usage du tabac confirment les données antérieures. Le risque associé au tabagisme est 3,8 fois plus élevé chez les gros fumeurs que chez les non-fumeurs, si l'on considère l'ensemble des sujets et le nombre total de cigarettes fumées.

Parmi les cas de cancer du pancréas, 75,4 % étaient des fumeurs comparative-ment à 63,6 % chez les témoins. Parmi les cas, 40,1 % étaient de gros fumeurs comparativement à 23,2 % chez les témoins [...].

Comme il a déjà été mentionné, 25 % des entrevues des cas ont été faites directement auprès des malades et 75 % d'entre elles auprès d'un proche. En ce qui concerne le nombre total de cigarettes fumées, le RC variait selon que l'entrevue avait été réalisée auprès des cas ou des proches. Dans le cas des entrevues directes, on a obtenu un RC de 3,04 pour tous les sujets (4,56 pour les hommes et 2,66 pour les femmes). Dans le cas des entrevues faites auprès d'un proche, le RC était de 5,35 pour tous les sujets (1,7 fois plus élevé qu'avec les entrevues directes) et de 6,1 pour les hommes (1,3 fois plus élevé). Le fait que les proches rapportent un risque plus élevé peut être dû à un biais de rappel concernant les habitudes de consommation de tabac des cas, comme cela a déjà été signalé.

L'ensemble des personnes qui buvaient de l'alcool présentait un risque moins élevé que l'ensemble des non-buveurs d'alcool (RC de 0,65) [...].

Les buveurs de café présentaient dans l'ensemble un moindre risque d'être atteints d'un cancer du pancréas que les non-buveurs de café (RC de 0,55) [...]. Bien qu'aucune étude n'ait, à notre connaissance, examiné la question des habitudes de consommation de café, nous avons trouvé que le risque, s'il y en a un, diminuait lorsque le café était consommé au cours d'un repas.

En fait, la seule conclusion livrée par les auteurs est que l'étude n'apporte rien de neuf par rapport à d'autres études antérieures, sauf peut-être cette diminution du risque lorsque le café est pris pendant les repas. La principale

EXEMPLE 14.3 (suite)

préoccupation des auteurs a été de souligner les lacunes que comporte la collecte de données effectuée auprès des proches. La discussion des autres biais et de la précision de l'étude est absente. En fait, la discussion se limite à un simple résumé des résultats déjà livrés dans la section « Résultats ». Elle peut donc être jugée comme étant inadéquate.

Outre la validité interne d'une étude, il importe également d'évaluer sa validité externe. Le concept de validité externe d'une étude renvoie au degré avec lequel ses conclusions demeurent valides lorsqu'elles sont appliquées à d'autres personnes en d'autres lieux. Dans l'exemple qui nous intéresse, il est difficile d'évaluer si les conclusions s'appliquent à la population de la polyclinique Milo, puisqu'on ne la connaît pas; la population à l'étude est mal définie.

Par ailleurs, les auteurs ne discutent pas de la signification clinique de leurs résultats. Toutefois, la conclusion principale, c'est-à-dire celle voulant que la consommation de tabac constitue un facteur de risque majeur pour le cancer du pancréas et que le risque augmente avec le nombre de cigarettes fumées, a des implications cliniques immédiates.

Autre information

Les auteurs mentionnent que l'étude fait partie du programme de recherche de l'Agence internationale de recherche sur le cancer de l'Organisation mondiale de la santé. Les informations sur une source de financement, sur une subvention pour la réalisation de l'étude ne sont pas présentées. On ne fournit pas non plus d'information sur l'approbation par un comité d'éthique ou de déclaration de conflit d'intérêts.

14.2.2 L'analyse critique d'une étude de cohorte(s)

Voici maintenant l'analyse critique de l'étude à visée étiologique de cohorte effectuée aux États-Unis par Zheng et ses collaborateurs entre 1966 et 1986 sur les facteurs de risque du cancer du pancréas. La grille d'analyse STROBE (*voir le tableau 14.1*) s'applique aussi aux études de cohorte(s) à la différence que pour certains éléments, on utilisera plutôt les énoncés propres à ce type d'étude. Par souci d'éviter la répétition des principes abordés précédemment, nous ferons preuve de concision.

Le titre et le résumé

Voici le résumé de l'étude de cohorte de Zheng et de ses collègues :

Les facteurs de risque du cancer du pancréas ont été évalués dans une cohorte de 17 633 Américains de race blanche qui ont répondu à un questionnaire posté en 1966 et qui ont été suivis jusqu'en 1986 pour ce qui est de la mortalité. Le tabagisme et la consommation d'alcool se sont révélés les facteurs de risque les plus importants pour le cancer du pancréas. Les risques augmentaient significativement en fonction du nombre quotidien de cigarettes fumées, allant jusqu'à être 4 fois plus élevés chez les fumeurs de 25 cigarettes ou plus que chez les non-fumeurs. L'alcool était également un facteur de risque significatif puisque les consommateurs de 10 portions ou plus par jour présentaient un risque 3 fois

plus élevé que les non-consommateurs d'alcool, mais la relation entre la quantité consommée et le risque n'était pas constante. La consommation de café n'était par ailleurs pas associée au risque. Par ailleurs, les analyses diététiques ont révélé une augmentation du taux de cancer du pancréas lorsque la consommation de viande était plus grande, même après ajustement selon les autres facteurs de risque. Les consommateurs de viande du dernier quartile risquaient environ trois fois plus de présenter un cancer du pancréas que ceux du premier quartile. Par contre, aucune association n'a été trouvée entre le cancer du pancréas et la consommation de fruits, de légumes ou de céréales. Cette étude confirme que la fumée de cigarette est un facteur de risque important pour le cancer du pancréas et apporte des preuves que l'alcool et la viande consommés en grandes quantités peuvent augmenter le risque de contracter cette néoplasie fatale.

Dans ce résumé, le but de l'étude, qui consiste à reconnaître les principaux facteurs de risque du cancer du pancréas, est clairement exposé. La méthode utilisée, à savoir l'observation sur une période de 20 ans de l'évolution de la mortalité dans une cohorte exposée à des facteurs de risque évalués en début d'étude à l'aide d'un questionnaire, est explicitement précisée. Les principaux résultats et les principales conclusions le sont aussi.

L'introduction

Le contexte/la justification et les objectifs Dans cet article, l'introduction présente une revue de la littérature sur le sujet. Comme le mentionnait l'étude cas-témoins, le tabagisme était alors le seul facteur de risque explicitement reconnu, en plus de l'histoire familiale. Les auteurs soulignent que le manque de crédibilité des études antérieures tient au fait qu'elles ont été pour la plupart rétrospectives et qu'une bonne partie de l'information obtenue l'a été auprès des proches des malades. Les chercheurs ne pouvaient dès lors être absolument certains qu'ils disposaient de données fiables, particulièrement en ce qui concernait l'alimentation. Pour pallier ce problème, Zheng et ses collègues proposent donc une étude de cohorte échelonnée sur une période de 20 ans. Ils se penchent sur les causes de mortalité chez 17 633 hommes de race blanche pour évaluer l'effet du tabac, de l'alcool et de l'alimentation sur le risque d'être atteint d'un cancer du pancréas.

On peut conclure que l'introduction présentée dans cet article satisfait à tous les critères de la grille d'évaluation. S'appuyant sur le bilan des connaissances du moment, on y présente le contexte et la légitimité de l'étude (item 2). On expose aussi clairement le but de l'étude (item 3).

La méthode

La conception de l'étude Les auteurs mentionnent qu'ils ont effectué une étude de cohorte. Le type d'étude est mentionné et bien décrit dans l'article.

Le contexte et la population La population à l'étude est décrite de la manière suivante dans cet article :

Cette étude de cohorte a été menée auprès des détenteurs d'une police d'assurance-vie de la Lutheran Brotherhood Insurance Society (LBS). En octobre 1966, un

questionnaire a été envoyé à 26 030 hommes blancs détenteurs d'une police. Ils avaient 35 ans ou plus et vivaient dans les États de la Californie, du New Jersey, de Washington, du Michigan, du Minnesota, du Missouri, du North Dakota, de l'Ohio et du Wisconsin. Au total, 17 818 hommes (68,5 %) ont rempli le questionnaire et ont été inclus dans la cohorte qui a été suivie pour voir l'évolution de la mortalité.

La population cible comprend les hommes blancs de plus de 34 ans. La population à l'étude est celle des hommes blancs de 35 ans ou plus qui possédaient une police d'assurance-vie de la LBS en 1966. Toute la population à l'étude a été invitée à participer à l'étude.

Disons d'abord que la grande force des études de cohorte(s) réside dans le fait qu'elles permettent de réduire le risque de biais de sélection. Un tel biais peut tout de même être introduit si le choix des sujets exposés et non exposés est influencé par la présence de la maladie étudiée. Ce problème peut survenir lorsqu'on étudie des maladies chroniques, des cancers par exemple, pour lesquelles la phase de latence est longue. Il devient alors possible de sélectionner un sujet porteur de la maladie, mais qui ne présente pas encore de symptômes. L'étude de Zheng et de ses collègues porte sur le cancer du pancréas et n'est pas à l'abri de cette lacune. Dans la section « Résultats », les chercheurs expliquent comment ils ont contrôlé ce biais potentiel :

L'exclusion des décès par cancer du pancréas survenus moins de cinq ans après le début du suivi n'a pas changé de façon significative les résultats relatifs aux effets de l'alimentation, de la consommation de café et d'alcool ou du tabagisme.

On voit que les auteurs, attentifs à cette erreur potentielle, mais impossible à prévenir, en tiennent compte au moment de l'analyse en démontrant son impact potentiel.

Les variables, les sources de données/mesures Dans cette étude, on a utilisé deux instruments. L'exposition au facteur étudié a été évaluée à l'aide d'un questionnaire, et l'évolution de la mortalité, grâce aux certificats de décès.

Les biais Les personnes qui ont été perdues de vue pendant le suivi sont à l'origine du biais de sélection le plus couramment rencontré dans les études de cohorte(s).

■ Non-réponse et refus de participer

Dans l'étude qui nous intéresse, 23 % des personnes ont été perdues de vue pendant le suivi. Les auteurs en discutent ainsi :

[...] environ un cinquième des membres de la cohorte (n = 4 027) a été perdu de vue pendant le suivi parce que la police d'assurance des détenteurs n'a pas été renouvelée ou est arrivée à échéance. Il n'existait pas de différence significative quant au tabagisme, à la consommation d'alcool ou à l'apport alimentaire entre les membres encore actifs après 20 ans de suivi et ceux qui ont été perdus de vue. En effet, une étude spéciale effectuée dans le but de connaître le statut vital de ceux qui avaient été perdus de vue après un suivi de 11,5 ans n'a pas permis de détecter de différence significative pour ce qui est de la mortalité par cancer dans les deux groupes. Il est donc peu probable que le suivi incomplet ait biaisé substantiellement nos résultats.

Dans l'étude de Zheng et de ses collègues, le tiers des personnes sollicitées pour l'étude a refusé d'y participer. Cependant, dans une étude de cohorte(s), cela ne pose pas un problème important. Dans une étude cas-témoins, le choix des sujets se fait en fonction du résultat final. La personne est malade ou non. C'est là la source de tous les biais. La façon de choisir les participants à l'étude aura nécessairement une influence sur la conclusion, puisque chaque personne choisie a été ou n'a pas été exposée au facteur étudié. Par contre, dans une étude de cohorte(s), les sujets sont choisis en fonction de l'exposition. Si aucune des personnes n'est malade au début du suivi, il ne peut y avoir de biais introduit par les personnes qui refusent de participer. Les problèmes liés au refus ne concernent pas la validité interne. Par contre, si le nombre de refus est grand, l'effectif de la cohorte peut devenir insuffisant pour assurer une précision et une puissance statistique acceptables. Le nombre de refus pourrait aussi influencer la validité externe de l'étude si l'on prétend que la cohorte est représentative de la population à l'étude. Cependant, même si Zheng et ses collègues ne discutent pas de ce problème, avec 17 000 participants, ou même 13 000 si l'on exclut les perdus de vue, suivis pendant une vingtaine d'années, cette étude conserve une bonne précision. Pour ce qui est de sa représentativité, nous en discuterons plus bas.

La validité du questionnaire Les auteurs ne disent rien des efforts déployés pour valider leur questionnaire. Il n'est pas possible à partir de l'information présentée par Zheng et ses collègues de connaître la précision du questionnaire utilisé pour évaluer l'exposition au début de l'étude.

On peut envisager qu'il s'agit là d'une source d'erreur systématique (non différentielle), mais on ne peut en évaluer l'ampleur. De même, on ne dispose pas d'information sur la validité des certificats de décès fournis par la compagnie d'assurances. Cependant, une erreur non différentielle, tout comme pour l'étude cas-témoins, ne pourra que sous-estimer l'effet réel (ce qui veut dire que si l'on trouve une association, elle sera réelle, mais simplement sous-estimée).

Une étude de cohorte(s) prospective est certes à l'abri de tous les biais liés à la mémoire. Mais lorsqu'on effectue un suivi différent selon l'exposition, on introduit forcément un biais différentiel. En principe, l'erreur différentielle peut tout aussi bien augmenter que diminuer la force de l'association. Souvent, on ignore quel est son effet. Ce genre de biais peut invalider complètement une étude. Dans l'étude de Zheng et de ses collègues, le problème ne se pose pas, puisque la seule intervention effectuée au cours du suivi est le constat de décès. Il n'y a pas de raison de penser que ce constat ait pu différer selon l'exposition. On introduira également un biais si l'on commet une erreur au moment de la classification des sujets en fonction de leur exposition. Ainsi, le fait de ne pas avoir suivi l'évolution de l'exposition tout au long des 20 ans peut avoir entraîné une erreur de classification. Comme le soulignent les auteurs eux-mêmes, si les fumeurs ont suivi la tendance des autres fumeurs américains au cours de cette période, plusieurs auront diminué leur consommation de tabac, voire auront cessé de fumer au cours de l'étude. Pourtant, pendant 20 ans, ils sont restés classés en fonction de leur exposition au début de l'étude, ce qui peut avoir créé une erreur différentielle. En effet, ce phénomène risque d'affecter beaucoup plus les sujets exposés qui ont davantage de chance d'avoir changé leurs habitudes que les

sujets non exposés. Dans un tel cas, il importe d'essayer de voir dans quel sens peut aller l'erreur. Ici, on suspecterait une sous-estimation du risque relatif puisque cela ferait diminuer artificiellement l'incidence du cancer chez les sujets exposés puisqu'ils fument moins. Le même raisonnement pourrait tenir pour les habitudes alimentaires s'il y a une amélioration de l'alimentation en vieillissant. Une autre erreur de classification différentielle possible dont les auteurs font état est la modification de l'alimentation des malades en raison précisément de leur maladie.

La taille de l'étude On ne mentionne pas les raisons qui ont conduit les chercheurs à solliciter 26 030 personnes, dont 17 818 ont pu être interrogées, plutôt que d'en solliciter 10 000 ou 100 000 pour constituer la cohorte. On comprend qu'il s'agissait là du nombre de clients de la compagnie d'assurances et qu'on a choisi ce nombre par commodité, mais on ignore quelle est la répercussion de ce choix sur la précision ou sur la puissance statistique de l'étude. Lorsqu'on étudie une maladie rare, la taille d'échantillon doit être grande pour permettre d'avoir la puissance requise. Zheng et ses collègues obtiennent ici des résultats statistiquement significatifs confirmant que la puissance est adéquate pour certaines associations, mais pas pour celle concernant le café (buveur de sept tasses de café et plus par jour comparativement à buveur de moins de trois tasses de café par jour : RR de 1,2 et ic de 0,3 à 2,4).

Les variables quantitatives et les analyses statistiques À la fin de la section «Méthode», un bref plan d'analyse est décrit, centré sur l'utilisation de méthodes statistiques pour contrôler le biais de confusion. De façon générale, on emploie les mêmes méthodes de contrôle du biais de confusion dans les études cas-témoins et les études de cohorte(s). Dans l'étude de Zheng et de ses collègues, on utilise une méthode de contrôle uniquement au moment de l'analyse (*a posteriori*), comme l'indique l'extrait suivant :

Le RR a été utilisé pour mesurer l'association entre la mortalité par cancer du pancréas et les facteurs d'exposition étudiés. On a appliqué la régression multiple de Poisson pour contrôler les facteurs de confusion possibles et pour le calcul des RR ajustés, avec un intervalle de confiance à 95 %.

Les principales variables de confusion potentielles connues à cette époque ont été incluses dans le modèle statistique et de façon globale. On peut donc aussi affirmer ici que les biais de confusion possibles ont été bien contrôlés.

Les résultats

La population L'information sur le nombre d'individus à chaque étape de l'étude a été présentée de façon adéquate à la section «Méthode».

Les données descriptives Les auteurs ne présentent pas dans un tableau, par exemple, les caractéristiques des groupes exposés et non exposés, ce qui rend difficile l'évaluation de la confusion potentielle.

Les données obtenues, les principaux résultats et les autres analyses Au début de l'étude, Zheng et ses collègues mesurent la consommation de tabac, d'alcool et de café chez les sujets de l'étude, puis ils mesurent la mortalité en cours de suivi à l'aide

de certificats de décès. Les auteurs présentent pour chaque facteur d'exposition étudié et pour différents niveaux d'exposition les nombres bruts de décès et le nombre de personnes-année, les risques relatifs bruts et ajustés et leur intervalle de confiance. Les mesures d'effet et les intervalles de confiance rapportés sont alors les suivants :

- gros fumeur comparativement à non-fumeur : RR de 3,9 et ic de 1,5 à 10,3 ;
- buveur d'alcool comparativement à non-buveur d'alcool : RR de 2,4 et ic de 1,1 à 5,5 ;
- buveur de 7 tasses de café et plus par jour comparativement à buveur de moins de 3 tasses de café par jour : RR de 1,2 et ic de 0,3 à 2,4 ;
- forte consommation de viande comparativement à faible consommation de viande : RR de 3,0 et ic de 1,2 à 7,5.

À première vue, quand on pense que le risque d'être atteint d'un cancer du pancréas chez un gros fumeur peut être de 1,5 à 10,3 fois plus élevé que chez un non-fumeur, il est vrai que cela n'est pas très précis.

Par ailleurs, les auteurs ont considéré dans leur analyse la possibilité de modification d'effet entre la consommation de viande, de tabac et d'alcool en utilisant la stratification.

La discussion et la conclusion

Les résultats clés et les limitations Les auteurs de l'étude de cohorte soulignent aussi que leurs conclusions vont dans le même sens que celles de la plupart des études antérieures.

Ce que nous avons observé dans notre cohorte va dans le même sens que les recherches antérieures qui indiquaient que le tabagisme constitue un risque majeur pour le cancer du pancréas. Le risque chez les fumeurs au moment de l'entrevue, en 1966, augmente en fonction du nombre de cigarettes fumées chaque jour. Il n'y a pas ou il y a peu d'excès de risque pour les ex-fumeurs, ce qui laisse croire que le fait d'arrêter de fumer peut diminuer le risque que ce cancer se développe [...]. Comme dans la plupart des études antérieures, nous n'avons pas trouvé d'association entre la consommation de café et le cancer du pancréas.

En ce qui concerne ce dernier point, les auteurs soulignent que la confirmation des conclusions des études antérieures étant obtenue ici dans le cadre d'une étude prospective, la preuve du caractère inoffensif du café est bien plus concluante. C'est ce qu'indique l'extrait suivant :

Cette observation, fondée sur les données d'une étude prospective, fournit des preuves supplémentaires que le café n'est pas un facteur de risque pour le cancer du pancréas.

L'alcool ressort toutefois nettement comme un facteur de risque, ce que ne démontreraient pas les études antérieures, en particulier les études cas-témoins. Dans ces études, le recours obligatoire à des proches pour la collecte de données induit un biais d'information non différentiel qui peut avoir en quelque sorte masqué l'association. Zheng et ses collègues soulignent aussi que cette preuve épidémiologique appuie ce

qu'on sait déjà de l'histoire naturelle de la maladie, c'est-à-dire que l'alcool est l'une des causes de pancréatite, et que la pancréatite prédispose au cancer.

La consommation d'alcool est significativement associée à la mortalité par cancer du pancréas dans notre étude, même si les degrés de consommation étaient relativement bas. Un excès de risque significatif de cancer du pancréas chez les gros buveurs a déjà été rapporté dans des études prospectives en Norvège et au Japon (pour les buveurs de whisky seulement). Par contre, la plupart des enquêtes réalisées auprès d'alcooliques n'ont pu établir un lien entre l'alcool et le cancer du pancréas, même si les auteurs d'une étude récente portant sur une cohorte d'alcooliques suédois ont noté un risque élevé de cancer du pancréas pour les deux sexes. Les résultats des études cas-témoins sont plus équivoques. Quelques études seulement ont établi un lien entre l'augmentation du risque et celle de la consommation d'alcool. Ces résultats contradictoires peuvent être dus à des différences dans la structure des études. De plus, dans la plupart de ces études, une grande proportion des cas étaient décédés, de telle sorte qu'on a dû recueillir les données auprès des proches. Ce recours aux proches a pu mener à une erreur de classification en ce qui a trait à l'exposition à l'alcool, et cette erreur, même si elle n'était probablement pas différentielle, peut avoir atténué l'association avec la maladie. Il est notoire que la consommation importante d'alcool peut provoquer une pancréatite chronique, un facteur de risque possible du cancer du pancréas.

Les auteurs discutent aussi longuement des facteurs diététiques considérés dans leur étude, celle-ci ayant démontré une relation significative entre la quantité de viande consommée et le cancer du pancréas. Par la suite, les auteurs soulignent avec beaucoup de pertinence que l'absence d'information sur les changements de régime alimentaire après le début de l'étude est susceptible d'avoir engendré un biais d'information non différentiel.

Les données concernant la diète n'ont été recueillies qu'une seule fois, en 1966. Or, plusieurs membres de la cohorte peuvent avoir changé leurs habitudes alimentaires au cours des 20 ans qu'a duré le suivi. Parce que cette erreur de classification est probablement non différentielle, elle aura tendance à ramener le RR vers 1 et à obscurcir une association faible ou même modérée avec certains facteurs diététiques.

Ils n'hésitent pas non plus à souligner d'autres erreurs possibles, notamment dans le passage suivant :

Notre étude a aussi d'autres limites. Comme dans le cas des informations sur la diète, nous n'avons aucune information supplémentaire sur l'usage du tabac après 1966, année où les sujets ont rempli le questionnaire. Si les fumeurs de cette cohorte ont suivi la tendance des hommes américains dans les dernières décennies, une proportion substantielle d'entre eux peut avoir arrêté de fumer pendant les 20 ans de suivi. Or, dans notre analyse, ils sont toujours considérés comme fumeurs, ce qui entraîne une sous-estimation de l'effet du tabagisme.

Les auteurs discutent aussi de la question des perdus de vue pendant le suivi et des travaux qu'ils ont faits pour en évaluer les répercussions, comme nous l'avons vu quand nous avons discuté des biais de sélection.

L'interprétation et la « généralisabilité » Les auteurs concluent que, malgré toutes ses lacunes, leur étude établit que le tabac constitue un risque majeur pour le cancer du pancréas. La consommation d'alcool et de viande semble augmenter le risque, mais les auteurs proposent que d'autres études soient réalisées pour établir plus clairement ce lien et les mécanismes biologiques sous-jacents. C'est ce qu'indique l'extrait qui suit :

En résumé, notre étude de cohorte confirme les découvertes des études antérieures, à savoir que l'habitude de fumer la cigarette constitue un facteur de risque majeur pour le cancer du pancréas. On a aussi observé que la consommation d'alcool et de viande augmentait le risque, bien que des études supplémentaires soient nécessaires pour clarifier ces associations et les mécanismes en jeu.

On peut donc affirmer que les auteurs de cette étude présentent une discussion adéquate de leurs résultats et fournissent une conclusion claire aux lecteurs. Bien qu'ils n'évoquent pas explicitement la signification clinique de leurs résultats, la conclusion principale concernant le lien entre la consommation de tabac et le cancer du pancréas a des implications cliniques immédiates. Elle renforce la détermination à travailler à des programmes d'arrêt du tabagisme, d'autant plus que, avec le temps, le risque pour les ex-fumeurs tend à diminuer pour se rapprocher de celui des non-fumeurs. Dans le cas de l'alcool et de la viande, le lien est moins clair. Un autre élément non négligeable qui ressort de l'étude, comme dans celle de Ghadirian et de ses collègues, c'est qu'on peut rassurer les patients buveurs de café.

Finalement, Lisa peut-elle conclure que ces résultats peuvent s'appliquer aux patients de la polyclinique Milo? L'étude de cohorte a été menée dans une population masculine américaine blanche âgée de plus de 34 ans et de revenu moyen (puisque les membres de la cohorte pouvaient se payer une police d'assurance). Elle conclut avec raison que les résultats de cette étude peuvent s'appliquer à une grande partie des patients qui consultent à la polyclinique Milo parce que ces derniers partagent les principales caractéristiques de cette population.

Par ailleurs, plusieurs autres études de cohorte en lien avec la consommation de café et le cancer du pancréas ont été publiées depuis les études décrites plus haut¹⁶. Celles-ci ont permis de mieux cerner les facteurs de risque du cancer du pancréas, qui sont principalement des susceptibilités génétiques, le tabagisme et l'obésité. L'alcool, certains éléments de la diète de même que certaines infections pourraient aussi être impliqués. Une étude de cohorte publiée en 2012¹⁷, incluant 229 119 hommes et 173 141 femmes américains recrutés en 1995 et suivis jusqu'en 2008 avec un suivi de 5 148 760 personnes-année, confirme que les buveurs de café peuvent être largement rassurés, car non seulement le café n'est pas en lien avec la mortalité par cancer, mais il semble même avoir un effet bénéfique sur une multitude de maladies comme le diabète, les maladies cardiovasculaires, les maladies respiratoires, etc.

Autre information Les informations sur une source de financement, sur une subvention, pour la réalisation de l'étude ne sont pas présentées. On ne fournit pas non plus d'information sur l'approbation de l'étude par un comité d'éthique. Il n'est donc pas possible de vérifier la présence de conflit d'intérêt chez les auteurs.

14.2.3 La recommandation clinique à la suite de l'analyse critique

Après cette rencontre fructueuse avec l'équipe de la polyclinique, Lisa revoit sa patiente Geneviève et peut avec raison lui mentionner que le tabagisme pourrait avoir contribué à l'apparition du cancer chez son père, mais que le café ne semble jouer aucun rôle.

Note sur l'évolution de la recherche avec les devis d'observation

Il est intéressant de constater qu'avec l'évolution des normes de conduite des études épidémiologiques, les différences, au chapitre des avantages et des limites, entre les études de cohorte(s) et cas-témoins, tendent à diminuer. En effet, de plus en plus d'études s'intéressant aux facteurs de risque des maladies utiliseront des bases de données administratives. Dans le cas du cancer du pancréas par exemple, l'issue pourra être documentée par l'utilisation de registres nationaux de cancer, complets et validés à l'échelle nationale. Lorsque l'exposition d'intérêt est disponible dans une base de données (registres d'assurance maladie, registre de médicaments, accès aux dossiers informatisés de médecins de première ligne, etc.), il devient possible d'étudier une population entière, sans biais de mémoire ou de rappel, avec des pertes de suivi minimales, en éliminant la longue période de suivi et en minimisant les coûts pour recueillir les données. Dans ces circonstances, la validité de l'étude revient presque entièrement à la fiabilité des bases de données utilisées. Une fois la population cible identifiée, elle sera essentiellement la même que la population à l'étude. Les groupes pourront être définis selon l'issue ou selon l'exposition. Selon l'issue (par exemple, l'identification de tous les cas de cancers par le registre des cancers, les autres personnes constituant le groupe des témoins) on pourra mener une étude cas-témoins en retraçant dans les registres l'exposition d'intérêt, en prenant en compte la période de latence présumée. Selon l'exposition, on pourra constituer une cohorte exposée et une autre non exposée pour ensuite vérifier l'incidence de la maladie dans chacun des groupes. Toutefois, l'investigation de l'effet de certaines expositions ne se prêtera jamais à l'étude par bases de données, comme celle des éléments de la diète ou de l'activité physique, puisque ces informations ne sont pas colligées systématiquement.

14.3 Les considérations éthiques

Des défis éthiques particuliers sont liés aux études expérimentales (ou essais cliniques randomisés), puisque dans ce type d'étude, c'est le chercheur qui détermine l'exposition. Il serait impensable d'exposer des sujets humains à des facteurs associés au développement de maladies dans le but de quantifier leur rôle. Pour cette raison, les devis non expérimentaux d'observation (études de cohorte(s) et cas-témoins) sont fréquemment employés en épidémiologie. Ces études sont du domaine de l'observation, car le recrutement des sujets se fait sur la base des caractéristiques observées (et non sur la base de caractéristiques imposées par le chercheur). Par exemple, on pourra recruter des fumeurs (exposition induite par le sujet lui-même) dans le cadre d'une étude de cohorte(s) pour étudier l'action délétère de la cigarette sur la santé.

Par contre, même si les études de cohorte(s) et cas-témoins relèvent du domaine de l'observation, il va sans dire qu'elles doivent aussi respecter des critères éthiques stricts reconnus internationalement tels que la justice, la bienfaisance et le respect (autonomie). Aucune considération de recherche ne pourrait justifier qu'un sujet reçoive, durant sa participation à un projet, des soins de qualité inférieure aux standards acceptés. Par exemple, au sein d'une étude de cohorte qui aurait pour but d'étudier le rôle des virus du papillome humain dans le développement du cancer du col, il serait inconcevable de laisser évoluer les lésions jusqu'au cancer, puisqu'il y a consensus sur le fait que les lésions précancéreuses graves doivent être traitées. La santé et la sécurité des personnes passent avant tout. Les comités d'éthique (et autres structures locales) des hôpitaux et des universités sont responsables du respect des règles éthiques reconnues. Leurs membres évaluent et approuvent les projets, y compris l'ensemble des procédures et des documents qui y sont liés. Le chercheur doit se conformer à ces règles et respecter les aspects éthiques dans toutes les dimensions de l'étude.

En résumé

Lorsqu'il est impossible de réaliser des études expérimentales ou quasi expérimentales pour explorer les facteurs de risque associés à certaines maladies, les études de cohorte(s) et les études cas-témoins sont des solutions de rechange fort acceptables. Les études de cohorte(s) présentent des risques moindres d'être entachées par des biais ou des erreurs, mais elles sont plus difficiles à mener sur le plan des conditions matérielles et financières. Les études cas-témoins sont moins onéreuses, mais le risque important qu'elles

comportent des biais contraint les chercheurs à une certaine prudence. Elles ne permettent généralement pas d'obtenir des conclusions définitives.

L'analyse critique des articles portant sur des études de cohorte(s) ou des études cas-témoins comprend une évaluation de leur validité interne, soit la recherche de biais de sélection, d'information et de confusion, ainsi qu'une appréciation de leur précision et, finalement, une évaluation de leur validité externe.

Exercices



Voir la section complète des exercices sur le Web.

1. Vrai ou faux.

- a) La meilleure approche pour rechercher les causes d'une nouvelle maladie fatale dont le taux d'incidence est de l'ordre de 1×10^{-5} an⁻¹ consiste à recruter une cohorte de plus de 5 000 personnes et à la suivre en évaluant les facteurs de risque les plus pertinents.
- b) Encore une fois, une nouvelle maladie frappe le pays. Elle n'est pas mortelle et guérit d'elle-même, mais elle oblige les patients à rester alités pendant 10 à 15 jours à cause d'étourdissements incoercibles. Il n'y a aucun traitement efficace et l'incidence de la maladie

est de l'ordre de 0,24 an⁻¹. Les études pathophysiologiques suggèrent une quinzaine de facteurs de risque plausibles. Dans ce cas, le suivi d'une cohorte de 100 personnes pendant 6 mois devrait permettre de trouver la cause de cette maladie.

2. Lequel des énoncés suivants désigne une caractéristique qui distingue les études cas-témoins des études longitudinales et transversales? Justifiez votre réponse.

- a) Les sujets peuvent être sélectionnés aléatoirement dans la population à l'étude.

- b) Les chercheurs peuvent constituer un ou plusieurs groupes témoins.
 - c) Il s'agit d'emblée d'études à visée étiologique.
 - d) Ce sont des études qui permettent de mesurer une association.
 - e) Ces études permettent d'éviter les biais de rappel.
- 3.** Parmi les énoncés suivants, lequel désigne une caractéristique distinctive des études de cohorte(s) non expérimentales ?
- a) Les sujets peuvent être sélectionnés aléatoirement dans la population à l'étude.
 - b) Les sujets sont classés au départ selon leur exposition à un ou plusieurs facteurs de risque.
 - c) Les chercheurs ne modifient pas les facteurs de risque déjà présents dans la population à l'étude.
 - d) Ce sont des études orientées dans le temps.
 - e) Ces études permettent d'éviter les biais de rappel.
- 4.** Dans une étude cas-témoins, certains biais peuvent survenir pour toutes les raisons évoquées ci-après, sauf une. Précisez laquelle, et justifiez votre réponse.
- a) Lors de l'entrevue, les interviewers peuvent rechercher inconsciemment les facteurs de risque chez les sujets qu'ils reconnaissent comme faisant partie du groupe des cas.
 - b) La connaissance des facteurs de risque présumés peut influencer le diagnostic.
 - c) Les sujets faisant partie du groupe des cas, de même que les témoins, peuvent avoir une mémoire défaillante.
 - d) Les sujets faisant partie du groupe des cas ont tendance à se souvenir de leur exposition passée aux facteurs de risque plus facilement que ne le font les témoins.
 - e) Le choix de témoins en milieu hospitalier peut parfois entraîner une sous-estimation du risque.

Pour les réponses, voir le corrigé à la fin du manuel.

La lecture critique des articles de synthèse systématique et des méta-analyses

Collaboration à la 3^e édition :

Agathe Croteau

Danielle Laurin

Au terme de ce chapitre, vous pourrez :

- expliquer les concepts et les méthodes de base des synthèses systématiques ;
- juger de la validité scientifique d'une synthèse systématique ;
- prendre une décision éclairée sur la pertinence de modifier votre pratique clinique à la lumière des résultats d'une synthèse systématique.

Objectifs spécifiques	Contenu
15.1 Expliquer l'utilité clinique des articles de synthèse systématique et en décrire les caractéristiques.	15.1 Les caractéristiques des différents types d'articles de synthèse
15.2 Décrire les avantages et les limites des articles de synthèse systématique.	15.2 Les avantages et les limites des articles de synthèse systématique
15.3 Décrire les principales étapes d'une synthèse systématique.	15.3 Les étapes de réalisation d'une synthèse systématique 15.3.1 L'interprétation des résultats
15.4 Procéder à l'analyse critique d'un article de synthèse systématique.	15.4 La critique des articles de synthèse systématique à l'aide de critères objectifs
15.5 Prendre en compte les résultats d'un article de synthèse systématique dans sa pratique.	15.5 La modification de sa pratique à la suite de la lecture d'un article de synthèse systématique
15.6 Décrire la contribution de la Collaboration Cochrane.	15.6 La Collaboration Cochrane





Mise en situation

Un groupe de médecins et de pharmaciens dont fait partie l'omnipraticienne Lisa s'interroge sur la pertinence d'utiliser une nouvelle catégorie de médicaments, les mimétiques de l'incrétine, pour le traitement du diabète de type 2. Ils se demandent s'il existe des preuves de l'efficacité et de l'innocuité de ce type de traitement. Autrement dit, ils veulent savoir si ces médicaments permettent d'améliorer la qualité des soins prodigués à leurs patients diabétiques.

Sur le moteur de recherche *PubMed*, Lisa effectue une recherche des articles publiés sur le sujet; elle en dénombre plus de 4000, dont près de 1400 articles de synthèse. Parcourant rapidement la liste des titres de ceux-ci, elle repère un article intitulé *Efficacy and safety of incretin therapy in type 2 diabetes: systematic review and meta-analysis*¹. Son choix se porte sur cet article, car il propose non seulement une analyse critique de plusieurs études, mais également une synthèse systématique et une méta-analyse. Bien que cet article date de quelques années, Lisa juge utile de prendre connaissance des études antérieures sur le sujet sans négliger pour autant les publications récentes. Elle souhaite cependant éviter aux membres du groupe ainsi qu'à elle-même la lecture d'un grand nombre d'articles. Elle leur propose donc de lire celui-ci et d'en discuter par la suite.

Pertinence clinique

Soucieux de sa formation continue, le professionnel de la santé préfère la lecture à toute autre méthode d'apprentissage. Mais devant l'abondance

des publications et les conclusions souvent contradictoires que l'on y trouve, il peine à utiliser efficacement son temps de lecture. Comme il ne peut lire un grand nombre d'études portant sur un même sujet pour en acquérir une connaissance approfondie, il choisit volontiers un texte qui résume et analyse les études récentes sur un sujet donné, soit l'article de synthèse (*review article*).

Certaines synthèses sont réalisées selon un protocole rigoureux; il s'agit de synthèses systématiques (*systematic review*). Assez souvent, la synthèse systématique comporte un volet quantitatif intégrant les résultats de plusieurs études, que l'on appelle «méta-analyse». Elle propose aussi, le cas échéant, une analyse des résultats contradictoires, utile à une prise de décision éclairée quant à la modification de la pratique clinique². Le professionnel de la santé doit être en mesure d'évaluer et de critiquer les articles de synthèse systématique.

Une pratique clinique s'appuyant sur des preuves scientifiques s'avère une compétence importante à acquérir pour les futurs diplômés dans les divers domaines de la santé, dont la médecine familiale³, et des services sociaux. Aussi, avant d'aller plus loin, il importe de définir le concept de «preuve scientifique». Il y a preuve scientifique lorsqu'il est possible d'établir un lien entre une exposition et un effet sur la santé.

Dans la hiérarchie de la force de la preuve selon le type d'étude consultée, les synthèses systématiques d'essais cliniques hasardisés figurent parmi les meilleures sources d'information lorsqu'il s'agit d'évaluer une intervention en santé⁴.

15.1 Les caractéristiques des différents types d'articles de synthèse

Il existe plusieurs types d'articles de synthèse dont la qualité et la valeur scientifique sont susceptibles de varier : les synthèses traditionnelles ou narratives et les synthèses systématiques avec ou sans méta-analyse. Fréquents dans la littérature, les éditoriaux, les consensus d'experts et les guides de pratique sont habituellement de forme narrative.

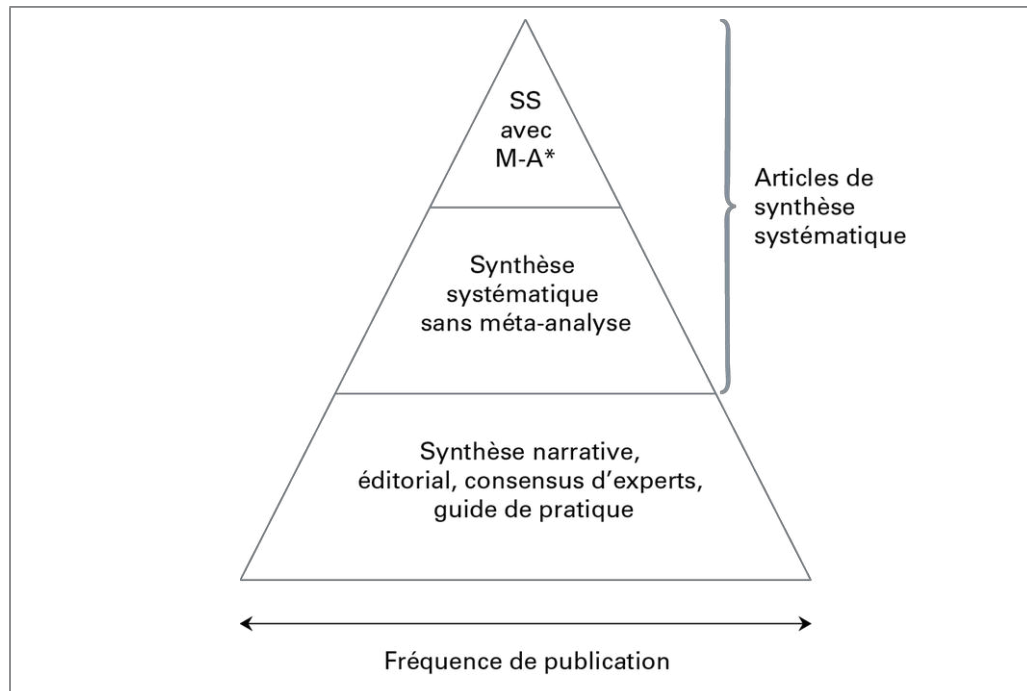
L'**éditorial** est un article écrit ou commandé par le comité de rédaction d'un périodique qui, en général, souhaite faire connaître son opinion sur un sujet. Les **consensus d'experts** sont des travaux de synthèse réalisés par des professionnels de la santé dont l'expertise est reconnue par leurs pairs. Les **guides de pratique** proprement dits sont rédigés par plusieurs experts qui formulent des recommandations à l'intention de leurs collègues après avoir procédé à une revue exhaustive de l'ensemble des preuves scientifiques sur un sujet particulier. Parfois, les conclusions tirées à partir de consensus d'experts font office de guide de pratique. Pour être qualifié d'**article de synthèse**, un éditorial, un consensus d'experts ou un guide de pratique doit aborder un seul sujet et s'appuyer sur des études publiant des résultats originaux.

Pour répondre à sa question de recherche, Lisa a trouvé un article de synthèse accompagné d'une méta-analyse. Elle sait que les articles de synthèse systématique constituent une forme particulière d'articles de synthèse pouvant comporter une méta-analyse. Soucieuse de bien utiliser ce type de publication scientifique, Lisa consulte quatre ouvrages de référence sur le sujet^{5, 6, 7, 8}.

Les articles de synthèse narrative ou systématique sont des articles scientifiques dans lesquels on résume l'état des connaissances sur une problématique donnée à partir des études originales. Les articles de synthèse systématique se distinguent des synthèses narratives par l'utilisation d'une méthode rigoureuse lors de la sélection et de l'analyse critique des articles retenus. De plus, ce type de synthèse comporte fréquemment des suggestions de pistes de recherche ou de développement concernant l'objet d'étude. La figure 15.1 présente une classification des articles de synthèse.

Lisa souhaite en apprendre davantage sur les articles de synthèse systématique comportant ou non une méta-analyse. Ses lectures révèlent que les synthèses systématiques sont réalisées selon un protocole préétabli et se caractérisent par :

- l'utilisation d'une stratégie de recherche visant à recenser toutes les études pertinentes sur un sujet précis ;
- des critères explicites servant à justifier l'inclusion et l'exclusion des études qui seront retenues ;
- des critères standardisés pour évaluer la qualité de ces études ;
- l'exploration des sources possibles d'**hétérogénéité** (variation parmi les résultats d'études) ;
- la synthèse quantitative de l'estimation de l'effet (dans le cas des méta-analyses) lorsque l'hétérogénéité est jugée acceptable.

FIGURE 15.1 ■ La classification et la fréquence des articles de synthèse

* Synthèse systématique avec méta-analyse

Parce que les articles de synthèse systématique répondent à des exigences méthodologiques bien définies et produisent de meilleures preuves scientifiques, la suite du chapitre portera sur les articles de synthèse systématique avec ou sans méta-analyse.

Les synthèses systématiques et les méta-analyses peuvent être réalisées à partir d'études expérimentales ou d'études d'observation. Cependant, puisque les études d'observation risquent davantage d'être affectées par des biais que les études expérimentales, la recherche des variations de l'estimation de l'effet selon les différentes sources de biais potentiels est toujours recommandée dans le cas d'une méta-analyse réalisée à partir d'études d'observation.

À retenir

- Les articles de synthèse résument les connaissances sur un sujet donné à partir d'études originales.
- Les articles de synthèse systématique, une forme particulière d'articles de synthèse, se caractérisent par la rigueur de leur processus de réalisation.
- Une synthèse systématique peut inclure une méta-analyse ou non.
- Une synthèse systématique peut être réalisée à partir d'études expérimentales, d'études d'observation ou à partir des deux.

15.2 Les avantages et les limites des articles de synthèse systématique

Pour le professionnel de la santé qui veut se maintenir à jour malgré l'abondance de connaissances et de publications dans son domaine, la lecture d'articles de synthèse s'avère essentielle. Elle lui permet d'épargner du temps et d'avoir accès à des commentaires critiques sur la validité des études recensées. Ces commentaires sont formulés par des experts et apparaissent d'emblée comme fondés. Cependant, les synthèses traditionnelles ou narratives sont souvent subjectives et sujettes à des biais ou à des faiblesses, telles l'inclusion sélective des études qui soutiennent l'opinion de l'auteur de la synthèse, la présentation de conclusions non basées sur une méthode explicite, une source d'information non spécifiée et l'absence d'évaluation standardisée de la qualité méthodologique des études. Contrairement aux synthèses traditionnelles narratives, les synthèses systématiques, avec ou sans méta-analyse, sont réalisées en suivant un protocole méthodologique rigoureux, qui améliore l'objectivité de leurs résultats.

Les auteurs des synthèses systématiques cherchent à recenser toutes les études pertinentes portant sur un sujet donné. L'évaluation de l'admissibilité et de la qualité méthodologique des études est justifiée par des critères explicites, et les résultats des études y sont rigoureusement analysés.

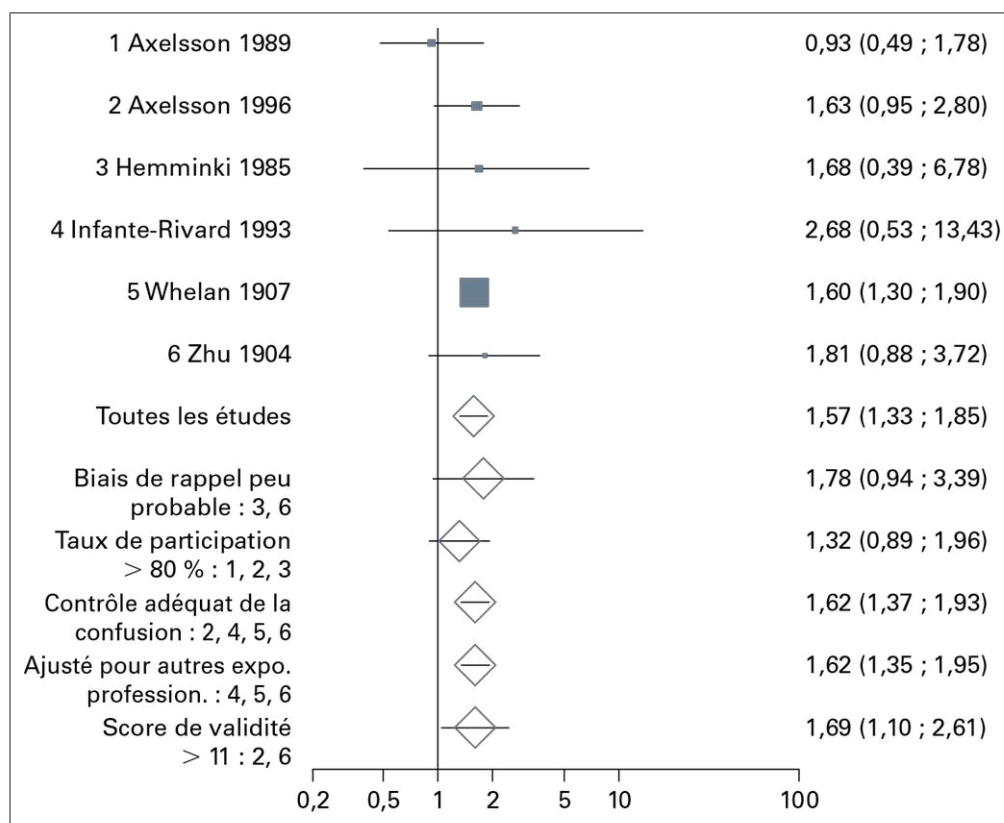
Dans un **article de synthèse systématique**, la méthode et les critères utilisés, présentés à la section « Méthode », permettent un processus de révision transparent et la reproductibilité d'une portion importante du travail de synthèse.

Dans une méta-analyse, la synthèse quantitative est obtenue par une moyenne pondérée des différents résultats, où les études ayant de grands effectifs ont plus d'influence ou de « poids ». Cette méthode a l'avantage de tenir compte des effectifs et des caractéristiques méthodologiques des études, contrairement à la pratique inadéquate qui consiste à dénombrer les études qui supportent une association *versus* celles qui ne la supportent pas.

EXEMPLE 15.1

Une méta-analyse a été réalisée afin de vérifier si le fait de travailler de nuit était associé au risque d'avortement spontané⁹. On présente les résultats de cette méta-analyse à la figure 15.2. Les six premières lignes illustrent les mesures d'association (RR ou RC) de chaque étude originale. Par exemple, la référence numéro 3 de la première ligne correspond au résultat publié par Axelsson et ses collaborateurs en 1989. À partir de la septième ligne, la mesure d'association synthèse et les résultats des analyses de sensibilité sont représentés par des losanges.

En observant les résultats des six études originales incluses dans cette méta-analyse, on remarque que cinq d'entre elles n'ont pu mesurer une association statistiquement significative. Par contre, la mesure d'association synthèse de toutes les études (septième ligne de la figure) montre une association statistiquement significative. Il est à noter que le poids relatif de chaque étude ayant servi au calcul de la moyenne pondérée est proportionnel à la taille du rectangle qui représente son résultat.

EXEMPLE 15.1 (suite)**FIGURE 15.2 ■ La présentation graphique des résultats: méta-analyse évaluant l'association entre le travail de nuit et le risque d'avortement spontané**

Source : Croteau, A. « L'horaire de travail et ses effets sur le résultat de la grossesse: méta-analyse et méta-régression: mise à jour », Institut national de santé publique du Québec, Montréal, 2007. Reproduit avec permission.

Ainsi, une synthèse quantitative peut améliorer la précision de l'estimation de la mesure d'association, mesurée dans cet exemple par le RR ou le RC. Il est à noter que le RC est une estimation juste du RR lorsque la fréquence de l'événement est rare (< 15%), ce qui était le cas pour les avortements spontanés dans l'exemple 15.1.

Cet exemple illustre l'amélioration de la puissance statistique que procurent les méta-analyses qui, en augmentant la taille des échantillons, accroissent la précision des résultats. Un autre avantage des méta-analyses, soit de chercher à expliquer l'hétérogénéité par l'évaluation de l'effet de certaines caractéristiques des études ou des sujets, fournit de l'information utile aux cliniciens. Les synthèses systématiques peuvent donc contribuer à dissiper l'incertitude lorsque les études originales, les articles de synthèse narrative et les éditoriaux sont en désaccord.

Malgré ces qualités, les synthèses systématiques n'énoncent pas de «vérité scientifique absolue»; leurs auteurs cherchent surtout à faire la synthèse, le plus objectivement possible, des données scientifiques disponibles sur un sujet donné. Pour atteindre cet objectif, ils utilisent une méthode systématique et standardisée pour la recherche des études et pour l'évaluation de la qualité des études retenues. L'emploi de grilles d'analyse et le recours à plusieurs personnes pour analyser les études leur permettent de poser un regard plus objectif sur les aspects méthodologiques de chacune des études sélectionnées. Cette forme d'analyse et l'établissement d'un processus destiné à résoudre les désaccords inter-observateurs ajoutent à la crédibilité de l'article de synthèse systématique.

Parce que la méthode et les critères décisionnels utilisés sont décrits dans les articles de synthèses systématiques, ces dernières, malgré une méthodologie rigoureuse, sont plus susceptibles d'être critiquées que les revues narratives ou les consensus d'experts qui ne suivent pas de méthode scientifique. Une synthèse systématique mal réalisée peut être biaisée en raison de l'exclusion d'études pertinentes, de l'inclusion d'études inadéquates ou d'une synthèse quantitative inappropriée. Par ailleurs, si la qualité méthodologique des études révisées est faible, les résultats de la synthèse laisseront à désirer. Enfin, un biais de publication peut altérer les résultats d'une synthèse systématique, parce que les études dont les résultats sont statistiquement significatifs ont plus de chance d'être publiées que les études dites négatives.

Cependant, l'application des principes à la base d'une bonne synthèse systématique et le respect des étapes de réalisation peuvent minimiser ces biais. Ainsi, les synthèses systématiques ont plus de chance d'éviter les biais que les synthèses narratives traditionnelles.

À retenir

- Les synthèses systématiques sont réalisées selon un protocole méthodologique rigoureux, ce qui améliore l'objectivité de leurs résultats.
- Les synthèses systématiques avec méta-analyse améliorent la précision de la mesure de l'estimation de l'effet et permettent l'exploration des sources d'hétérogénéité.

15.3 Les étapes de réalisation d'une synthèse systématique

La réalisation d'une synthèse systématique suit un processus scientifique qui exige autant de minutie qu'une étude originale expérimentale ou d'observation. On précise dans un protocole les objectifs, les hypothèses et la méthode. Les étapes de réalisation d'une synthèse systématique, incluant une méta-analyse, sont les suivantes.

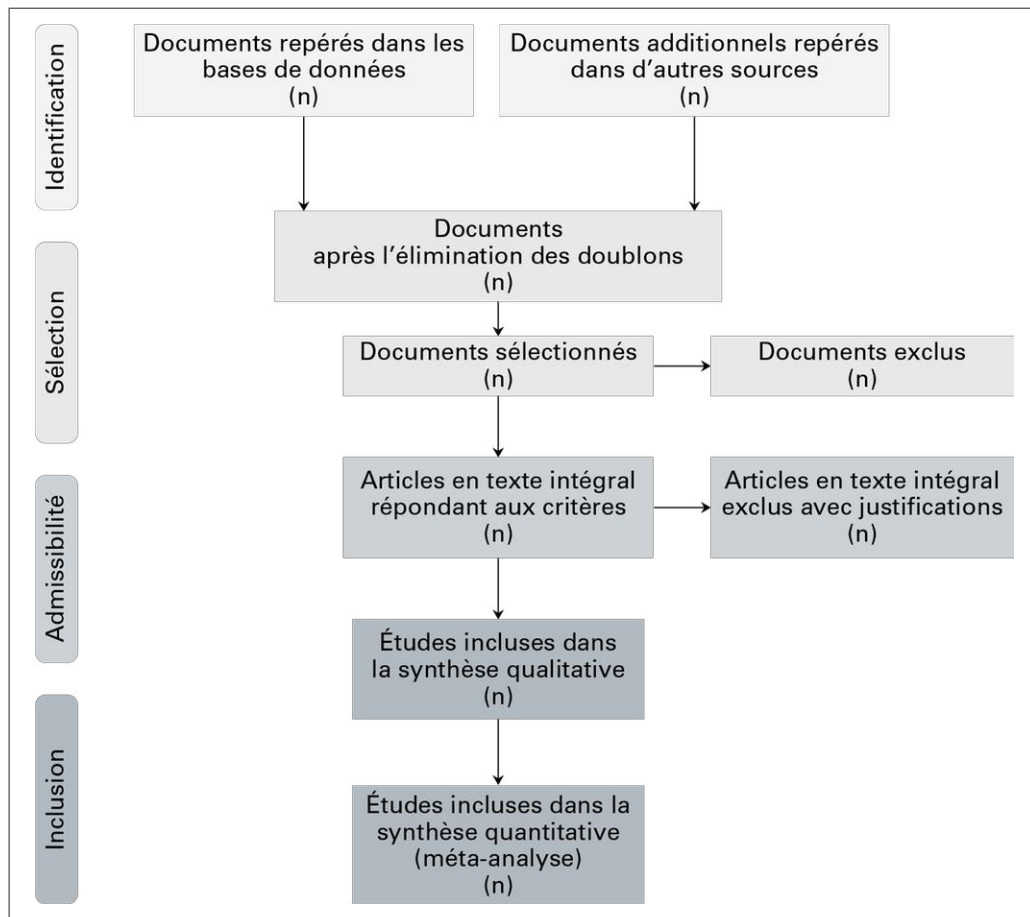
1. Formulation de la question de recherche.
2. Définition des critères d'inclusion et d'exclusion des études (population, intervention ou exposition, effet sur la santé, devis des études).
3. Recherche des publications (stratégie de recherche, bases de données interrogées, autres sources de données).
4. Sélection des études, idéalement par plus d'un observateur (arbitrage ou consensus en cas de désaccord, tenue d'un registre des études incluses et exclues et des motifs d'exclusion).
5. Évaluation de la qualité méthodologique des études, idéalement par plus d'un observateur (stratégie en cas de désaccord, utilisation d'une liste de critères).
6. Extraction des données, idéalement par plus d'un observateur (stratégie en cas de désaccord, définition des variables relatives aux effets sur la santé, à l'exposition ou au traitement, et définition des variables de confusion).
7. Analyse et présentation des résultats (tableaux présentant les caractéristiques et les résultats des études individuelles, exploration des sources possibles d'hétérogénéité par des analyses de sensibilité, de sous-groupes* ou par **méta-régression***, jugement sur la pertinence d'effectuer une synthèse quantitative, calcul des mesures synthèses* soit pour l'ensemble des études, soit par sous-groupes d'études ou en incluant les deux, graphiques de type «forest»*, recherche d'un biais de publication).
8. Interprétation des résultats (considération des biais possibles incluant le biais de publication, discussion concernant les limites de la synthèse systématique, considération de l'applicabilité des résultats, évaluation de l'importance relative des effets positifs *versus* les effets indésirables, suggestion de futures avenues de recherche).

Avant d'aller plus loin, il importe de préciser et d'illustrer certaines étapes propres à la synthèse systématique.

Le processus de recherche des publications et de sélection des études est souvent présenté sous la forme d'un **diagramme de flux** (*flow diagram*) qui dénombre les études recensées, incluses et exclues, ainsi que les motifs d'exclusion.

Un groupe d'experts de différentes disciplines a produit l'énoncé PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), un outil visant à améliorer la qualité des publications d'articles de synthèse systématique et de méta-analyse¹⁰. L'énoncé PRISMA comporte une liste de critères et un diagramme de flux. Le diagramme de flux présenté à la figure 15.3 illustre comment rapporter le processus de sélection des études pour une synthèse systématique. Notez que les parenthèses sous chaque énoncé servent à indiquer le nombre de documents correspondant.

* Éléments de traitement quantitatif et statistique des résultats qui s'appliquent exclusivement à la méta-analyse.

FIGURE 15.3 ■ Le diagramme de flux de l'énoncé PRISMA

Source: Moher, D., Liberati, A., Tetzlaff, J. et autres. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*, 2009, vol. 6, n° 7, e1000097. doi: 10.1371/journal.pmed.1000097. Traduction française réalisée par la Bibliothèque du Centre hospitalier de l'Université de Montréal (CHUM) (2014).

Dans une méta-analyse, on utilise des méthodes statistiques pour obtenir une mesure d'association synthèse qui correspond à une moyenne pondérée des résultats de chaque étude originale. Le plus souvent, les rapports d'études originales présentent leurs résultats en indiquant le nombre de sujets dans chaque groupe (par exemple, traités/non traités, malades/non malades) ou encore par des mesures d'association (par exemple, RR ou RC). Dans certains cas, les résultats à combiner peuvent être des risques attribuables, des rapports de taux, des différences de moyennes, etc. D'ailleurs, des méthodes statistiques permettent d'en faire la synthèse (voir Egger² et Sutton⁸). Si des résultats de nature différente ne peuvent être combinés, les RR et les RC peuvent l'être si la fréquence de l'événement est rare (< 15 %).

Deux modèles statistiques permettent d'obtenir des mesures d'association synthèse : le modèle à effets fixes et le modèle à effets aléatoires. Le modèle à effets fixes pondère les résultats d'études selon la quantité d'information qu'elles contiennent. L'effectif

d'une étude est le principal déterminant de la quantité d'information qu'elle comporte. Cependant, cette méthode ne tient pas compte de la variation des résultats entre les études (hétérogénéité). Le modèle à effets aléatoires permet d'ajouter une quantité de variances dues à l'hétérogénéité dans le calcul de la mesure synthèse. En présence d'hétérogénéité, il produit généralement un résultat moins sensible aux effectifs des études avec un intervalle de confiance plus large autour de la mesure synthèse. Pour ce modèle, la méthode statistique la plus simple et la plus utilisée est celle de DerSimonian et Laird^{2, 8, 11}. Les résultats présentés à la figure 15.2 ont été obtenus par cette méthode.

L'évaluation statistique du degré d'hétérogénéité et l'exploration des causes possibles de l'hétérogénéité sont des étapes importantes d'une méta-analyse^{2, 8, 12}. Les méthodes statistiques basées sur le modèle à effets aléatoires permettent de tenir compte de l'hétérogénéité, mais ne peuvent l'expliquer. Les analyses de sensibilité, les **analyses par sous-groupes** ou les méta-régressions, décrites plus loin, sont utiles pour explorer les sources d'hétérogénéité. Si, par la suite, une quantité importante d'hétérogénéité demeure inexplicée, un jugement doit être porté pour décider s'il est encore approprié de combiner les résultats, et la priorité devrait être d'essayer de comprendre les causes de l'hétérogénéité^{2, 7, 8, 13}. Il n'existe aucune règle claire indiquant à partir de quel degré d'hétérogénéité (voir le tableau 15.1) les résultats des études ne devraient pas être combinés. La décision de combiner les résultats ne devrait pas reposer seulement sur des considérations statistiques, mais aussi sur la compréhension des éléments scientifiques et contextuels liés à la question de recherche.

TABLEAU 15.1 ■ L'évaluation statistique de l'hétérogénéité¹²

Khi carré (χ^2) d'hétérogénéité	Permet de vérifier si les différences observées parmi les résultats sont compatibles avec le seul effet du hasard.	Une valeur $p < 0,10$ indique la présence d'hétérogénéité.
I^2	Exprime la proportion de la variabilité totale qui est attribuable à l'hétérogénéité plutôt qu'au hasard.	0 – 40% : hétérogénéité non importante 30 – 60% : hétérogénéité modérée 50 – 90% : hétérogénéité substantielle 75 – 100% : hétérogénéité considérable

Les analyses de sensibilité et les analyses par sous-groupes s'avèrent utiles pour explorer les sources d'hétérogénéité telles que certaines caractéristiques des patients et des études. Les analyses par sous-groupes indiquent si les résultats diffèrent selon les strates de variables telles que l'âge, le sexe, l'état de santé au départ et les divers traitements évalués. Par ailleurs, les analyses de sensibilité présentent des mesures synthèses pour des sous-ensembles d'études définis par des critères d'inclusion plus sévères (par exemple, un suivi plus long, à double insu, le masquage de l'allocation, une faible proportion de sujets perdus au suivi). Ces caractéristiques des patients ou des études sont définies au préalable selon des considérations théoriques. Dans le contexte d'études d'observation, les analyses de sensibilité sont particulièrement utiles pour vérifier la similitude des résultats selon des caractéristiques méthodologiques liées à la validité des études afin d'explorer les effets de biais potentiels.

L'évaluation approfondie des effets de ces biais peut apporter un éclairage supérieur à celui d'une synthèse quantitative globale de l'ensemble des études.

EXEMPLE 15.2

Dans la méta-analyse d'études d'observation dont les résultats sont présentés à la figure 15.2, différentes analyses de sensibilité ont été effectuées. Les résultats de ces analyses, réalisées pour différents sous-ensembles d'études qui présentaient de bonnes caractéristiques méthodologiques, apparaissent aux cinq dernières lignes de cette figure. L'auteure a voulu vérifier la similitude des résultats obtenus par les analyses de sensibilité, en les comparant au résultat global obtenu pour l'ensemble des études. Dans l'exemple, les résultats de quatre études ont été ajustés pour les variables personnelles potentiellement confondantes. Le résultat de cette **analyse de sensibilité** ne comprenant que quatre études, « Contrôle adéquat de la confusion », montre peu de variation comparativement au résultat de la synthèse quantitative globale des six études de « Toutes les études ».

La méta-régression, une autre approche pour expliquer l'hétérogénéité, permet de vérifier si l'hétérogénéité peut s'expliquer par un ou plusieurs facteurs. Dans une méta-régression, chaque étude est considérée comme un « sujet », et l'on traite chacune des caractéristiques des études comme une variable du modèle. Le principal avantage de la méta-régression est qu'elle tient compte de plusieurs caractéristiques simultanément. Cependant, moins il y a d'études et plus le nombre de caractéristiques à prendre en compte est grand, moins la méta-régression sera utilisable.

15.3.1 L'interprétation des résultats

En général, on présente les résultats des méta-analyses (synthèse quantitative globale, par sous-groupes et analyses de sensibilité) sous forme de graphiques de type « forest ». La figure 15.2 en est un exemple.

Les auteurs d'articles de synthèse s'exposent à un risque, soit la possibilité d'un biais de publication ; les études ayant obtenu des résultats statistiquement significatifs ont plus de chance d'être publiées que les études n'atteignant pas le seuil de signification statistique (voir « *Le biais de publication* » dans le chapitre 6, à la page 115). Un tel biais peut sérieusement invalider les résultats d'une synthèse systématique, d'où l'importance de chercher à inclure toutes les études pertinentes, publiées dans leur intégralité (par exemple, l'article original) ou non (par exemple, les résultats préliminaires présentés dans le cadre d'un congrès sous forme de résumé, une thèse, un rapport).

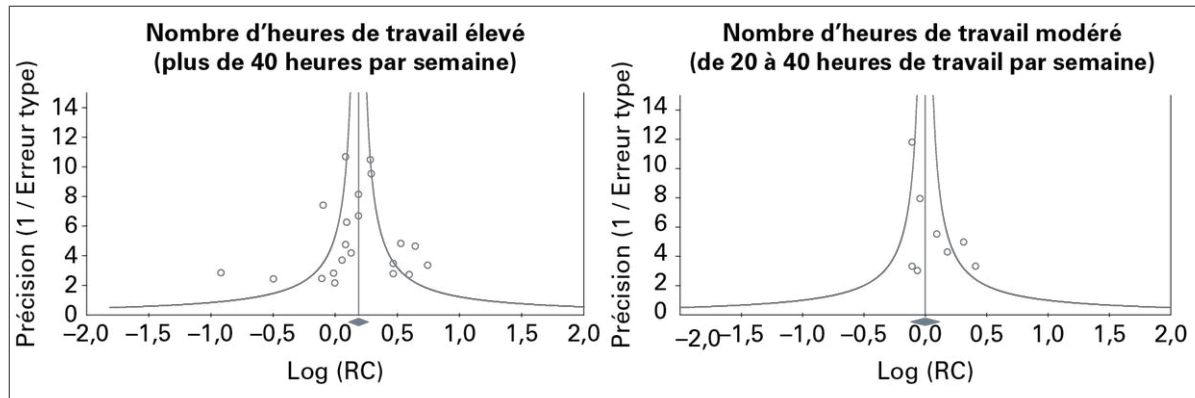
Quelques méthodes statistiques et graphiques permettent d'évaluer la plausibilité d'un biais de publication. Cependant, aucune ne donne de certitude à propos de l'existence ou de l'inexistence d'un biais de publication. Une des méthodes les plus faciles à utiliser fait appel à un diagramme de dispersion (*funnel plot*). L'échelle de l'ordonnée (axe des y) de ce graphique correspond à la précision (par exemple, l'inverse de l'erreur type) de la mesure d'association ; les points représentant les résultats des études de plus faible effectif seront donc situés vers le bas. La mesure d'association se trouve sur l'axe des x ; on utilise une échelle logarithmique pour des mesures d'association comportant un rapport de mesures (par exemple, RR, RC). En l'absence de biais de publication, on peut

s'attendre à dénombrer environ autant de petites études de chaque côté de la mesure d'association synthèse (fortement influencée par les grandes études). Dans ce cas, la répartition symétrique des points au bas du graphique permet d'estimer improbable le biais de publication. En revanche, en présence d'un biais de publication, des études n'ayant pas obtenu un résultat statistiquement significatif, en général plus petites, seront absentes. Il y aura donc moins de points d'un côté de la mesure d'association synthèse. Une répartition asymétrique suggère un biais de publication^{2, 8, 12}.

EXEMPLE 15.3

Afin de vérifier la plausibilité d'un biais de publication parmi des études ayant évalué l'effet des heures de travail hebdomadaire sur le risque d'accouchement avant terme, deux diagrammes de dispersion, présentés à la figure 15.4, ont été produits.

FIGURE 15.4 ■ Des diagrammes de dispersion des études sur l'association entre le nombre d'heures de travail élevé ou modéré et le risque d'accouchement avant terme



Le graphique de gauche (> 40 heures de travail) démontre une répartition assez symétrique des points; on estime le biais de publication improbable. Par contre, dans le graphique de droite, la répartition asymétrique, soit un plus grand nombre de points à droite de la mesure d'association synthèse, laisse supposer que de petites études dont les résultats sont négatifs n'ont pas été publiées.

Une meilleure connaissance des articles de synthèse systématique et de leurs limites permettra au lecteur de poser un regard critique sur leurs résultats.

À retenir

- La sélection des études se fait à partir de critères d'inclusion et d'exclusion; cette démarche est décrite à l'aide d'un diagramme de flux.
- La qualité méthodologique de chaque étude est évaluée à l'aide de critères précis.
- L'analyse des résultats comporte une évaluation de l'hétérogénéité par des méthodes statistiques et la recherche des sources d'hétérogénéité.
- L'interprétation des résultats tient compte des biais possibles.

15.4 La critique des articles de synthèse systématique à l'aide de critères objectifs

Dans un premier temps, les membres du groupe de médecins et de pharmaciens de la polyclinique Milo font une critique de l'article sélectionné par Lisa. Dans un second temps, ils entendent déterminer dans quelle mesure cet article pourra répondre à leurs interrogations. De plus, ils se proposent de vérifier si les conclusions avancées par les auteurs sont suffisamment probantes (c'est-à-dire de qualité scientifique élevée) pour en tenir compte dans leur pratique clinique.

À partir d'outils existants, des experts ont élaboré une grille d'évaluation de la qualité méthodologique des revues systématiques: AMSTAR¹⁴ (*a measurement tool to assess the methodological quality of systematic reviews*). Par la suite, des critères d'attribution d'un score y ont été ajoutés pour obtenir la liste de contrôle AMSTAR révisée: R-AMSTAR¹⁵. Ces outils ont été traduits en français par l'Institut national d'excellence en santé et en service sociaux (INESSS)¹⁶.

Marie-Louise, la pharmacienne du groupe, propose d'utiliser la liste de contrôle R-AMSTAR, dont les 11 questions sont présentées dans l'encadré 15.1.

ENCADRÉ 15.1 ■ Les questions de la liste de contrôle R-AMSTAR^{15, 16}

1. Un plan de recherche établi *a priori* est-il fourni?

La question de recherche et les critères d'inclusion des études doivent être déterminés avant le début de la revue.

2. La sélection des études et l'extraction des données ont-elles été confiées à au moins deux personnes?

Au moins deux personnes doivent procéder à l'extraction des données de façon indépendante, et une méthode de consensus doit avoir été mise en place pour le règlement des différends.

3. La recherche documentaire était-elle exhaustive?

Au moins deux sources électroniques doivent avoir été utilisées. Le rapport doit comprendre l'horizon temporel de la recherche et les bases de données interrogées (Central, EMBASE et MEDLINE, par exemple). Les mots-clés et (ou) les termes MeSH doivent être indiqués et, si possible, la stratégie de recherche complète doit être exposée. Toutes les recherches doivent être complétées par la consultation des tables des matières de revues scientifiques récentes, de revues de la littérature, de manuels, de registres spécialisés ou d'experts dans le domaine étudié et par l'examen des références fournies dans les études répertoriées.

4. La nature de la publication (littérature grise*, par exemple), était-elle un critère d'inclusion?

Les auteurs doivent indiquer s'ils ont recherché tous les rapports, quel que soit le type de publication, ou s'ils ont exclu des rapports (de leur synthèse systématique) sur la base du type de publication, de la langue, etc.

* La littérature grise correspond à tout document produit par les organismes gouvernementaux ou universitaires, les entreprises et l'industrie, sans contrôle par l'édition commerciale.

ENCADRÉ 15.1 ■ Les questions de la liste de contrôle R-AMSTAR^{15, 16} (suite)

5. Une liste des études (incluses et exclues) est-elle fournie ?

Une liste des études incluses et exclues doit être fournie.

6. Les caractéristiques des études incluses sont-elles indiquées ?

Les données portant sur les sujets qui ont participé aux études originales, les interventions qu'ils ont reçues et les résultats doivent être regroupés, sous forme de tableau, par exemple. La distribution des données sur les caractéristiques des sujets de toutes les études analysées doit y figurer.

7. La qualité scientifique des études incluses a-t-elle été évaluée et consignée ?

Les méthodes d'évaluation déterminées *a priori* doivent être indiquées (par exemple, pour les études sur l'efficacité pratique, le choix de n'inclure que les essais cliniques randomisés à double insu avec placebo ou de n'inclure que les études où l'affectation des sujets aux groupes d'étude était dissimulée); pour d'autres types d'études, d'autres critères d'évaluation seront à prendre en considération.

8. La qualité scientifique des études incluses dans la revue a-t-elle été utilisée adéquatement dans la formulation des conclusions ?

Les résultats de l'évaluation de la rigueur méthodologique et de la qualité scientifique des études incluses doivent être pris en considération dans l'analyse et les conclusions de la revue, et formulés explicitement dans les recommandations.

9. Les méthodes utilisées pour combiner les résultats des études sont-elles appropriées ?

Si l'on veut regrouper les résultats des études, il faut effectuer un test d'homogénéité afin de s'assurer qu'elles sont combinables (khi carré ou I^2 , par exemple). S'il y a hétérogénéité, il faut utiliser un modèle d'effets aléatoires et (ou) vérifier si la nature des données cliniques justifie la combinaison (la combinaison est-elle raisonnable?).

10. La probabilité d'un biais de publication a-t-elle été évaluée ?

Une évaluation du biais de publication doit comprendre une association d'outils graphiques (diagramme de dispersion, des études ou autre test) et (ou) des tests statistiques (test de régression d'egger, par exemple).

11. Les conflits d'intérêts ont-ils été déclarés ?

Les sources possibles de soutien doivent être déclarées, tant pour la revue systématique que pour les études qui y sont incluses.

Dans le texte qui suit, nous passerons en revue chacune des questions de la grille d'évaluation de la qualité méthodologique des revues systématiques AMSTAR et nous appliquerons les critères d'attribution du score R-AMSTAR, en examinant l'article que Lisa et ses collègues entendent critiquer.

1. Un plan de recherche établi *a priori* est-il fourni ?

La question de recherche et les critères d'inclusion des études doivent être déterminés avant le début de la revue.

- Critère A : Publication et (ou) inscription du protocole d'étude à l'avance (Il doit être dit explicitement que le protocole a été inscrit ou publié dans un registre de revues systématiques, par exemple dans PROSPERO, un registre de revues systématiques prospectif et multinational accessible en ligne.)
- Critère B : Description des critères d'inclusion
- Critère C : Question de recherche bien ciblée (La question renferme les critères «PICO», soit Population, Intervention [ou exposition], Comparateur [ou témoins] et Résultats [ou *Outcomes*].)
- Conditions d'attribution du score : 3 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4

Marie-Louise rapporte qu'elle n'a pu trouver d'information concernant la publication ou l'inscription à l'avance du protocole de la synthèse systématique.

Dans la section « Méthode » de l'article, on mentionne les critères d'inclusion suivants :

[...] données originales tirées d'essais cliniques hasardisés, d'une durée de 12 semaines et plus, réalisés dans une population d'adultes (femmes non enceintes) atteints de diabète de type 2, chez laquelle l'effet sur le niveau d'hémoglobine glycosylée (HbA_{1c}) a été comparé entre des sujets traités par un mimétique de l'incrétine et des sujets traités par un hypoglycémiant ou recevant un placebo.

Concernant la question de recherche, on peut lire ceci à la fin de l'introduction de l'article :

Cette méta-analyse évalue l'efficacité et l'innocuité des mimétiques de l'incrétine (agonistes du glucagon-like peptide 1 [GLP-1] et inhibiteurs de la dipeptidyl-peptidase-4 [DPP4] chez des adultes (femmes non enceintes) atteints de diabète de type 2, à partir d'essais cliniques hasardisés publiés et non publiés.

Les critères B et C sont respectés ; le score est de 3 points sur 4.

2. La sélection des études et l'extraction des données ont-elles été confiées à au moins deux personnes ?

Au moins deux personnes doivent procéder à l'extraction des données de façon indépendante, et une méthode de consensus doit avoir été mise en place pour le règlement des différends.

- Critère A : Données extraites par au moins deux personnes, de façon indépendante (déclaration explicite ou implicite)
- Critère B : Énoncé sur le processus de consensus pour le règlement des différends

- Critère C : Résolution des désaccords entre les personnes ayant extrait les données conformément à la méthode établie (déclaration explicite ou implicite)
- Conditions d'attribution du score : 3 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4 (suite)

À la rubrique «Extraction des données» du résumé, on peut lire :

Deux réviseurs ont évalué les études de façon indépendante, ont établi celles à inclure dans la méta-analyse et en ont extrait des données.

Après la lecture de la section «Méthode», on retient ceci :

Deux auteurs de l'article (R.E.A et A.G.P.) ont révisé indépendamment les résumés des études conformément aux critères d'inclusion. De plus, les désaccords étaient réglés à la suite d'un consensus entre ces deux réviseurs indépendants, ou en conférence de groupe à partir de l'article original.

Les trois critères sont respectés ; le score est de 4 points sur 4.

3. La recherche documentaire était-elle exhaustive ?

Au moins deux sources électroniques doivent avoir été utilisées. Le rapport doit comprendre l'horizon temporel de la recherche et les bases de données interrogées (Central, EMBASE et MEDLINE, par exemple). Les mots clés et (ou) les termes MeSH doivent être indiqués, et, si possible, la stratégie de recherche complète doit être exposée. Toutes les recherches doivent être complétées par la consultation des tables des matières de revues scientifiques récentes, de revues de la littérature, de manuels, de registres spécialisés ou d'experts dans le domaine étudié et par l'examen des références fournies dans les études répertoriées.

- Critère A : Au moins deux sources électroniques ont été utilisées.
- Critère B : L'horizon temporel et les bases de données interrogées sont indiqués.
- Critère C : Les mots clés et (ou) les termes MeSH sont indiqués, et, si possible, la stratégie de recherche est exposée.
- Critère D : Toutes les recherches sont complétées par la consultation des tables des matières de revues scientifiques récentes, de revues de la littérature, de manuels ainsi que de registres et par l'examen des références fournies dans les études répertoriées.
- Critère E : Une recherche manuelle a été effectuée dans les revues (la recherche manuelle consiste à repérer les revues très pertinentes et à faire une recherche à la main, page par page, de leur contenu afin de relever d'éventuelles études admissibles).
- Conditions d'attribution du score : 4 ou 5 critères → 4 points, 3 → 3, 2 → 2, 1 ou 0 → 1

4. La nature de la publication (littérature grise, par exemple) était-elle un critère d'inclusion?

Les auteurs doivent indiquer s'ils ont cherché tous les rapports, quel que soit le type de publication, ou s'ils ont exclu des rapports (de leur synthèse systématique) sur la base du type de publication, de la langue, etc.

- Critère A: Les auteurs indiquent qu'ils ont recherché tous les rapports, quel que soit le type de publication.
- Critère B: Les auteurs indiquent s'ils ont exclu des rapports sur la base du type de publication, de la langue, etc.
- Critère C: Les articles rédigés dans une langue autre que l'anglais ont été traduits ou les lecteurs maîtrisaient assez bien la langue du rapport.
- Critère D: Aucune restriction fondée sur la langue ou prise en compte des articles rédigés dans une langue autre que l'anglais
- Conditions d'attribution du score: 3 ou 4 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4 (suite)

Pour les questions 3 et 4, dans la section « Méthode » de l'article, on mentionne que :

Les bases de données MEDLINE (entre 1966 et le 20 mai 2007) et Cochrane (registre des essais cliniques hasardisés, second trimestre 2007) ont été interrogées afin de recenser des rapports, en langue anglaise, d'essais cliniques hasardisés à partir d'une liste des mots clés, mentionnée dans l'article. La recherche a été complétée en consultant la documentation concernant les médicaments approuvés, deux sites Internet pertinents, les références personnelles et la bibliographie des articles retracés. Des résumés d'essais cliniques, non publiés dans les périodiques avec comité de pairs, mais présentés à des conférences internationales ont aussi été ajoutés; les auteurs de la synthèse systématique ont justifié cet ajout.

Pour la question 3, les critères A, B et C sont respectés, et le critère D est partiellement respecté; le score est de 3 points sur 4.

Pour la question 4, puisque des essais cliniques non publiés dans les périodiques mais présentés en conférence ont été ajoutés, les critères A et B sont respectés; le score est de 3 points sur 4.

5. Une liste des études (incluses et exclues) est-elle fournie?

Une liste des études incluses et exclues doit être fournie. Les études exclues sont celles qui, après avoir été sérieusement prises en considération sur la foi du titre et (ou) du résumé, ont été rejetées après lecture du corps du texte.

- Critère A: Les études incluses doivent être réunies dans un tableau, une liste ou une figure; une simple liste de références ne suffit pas.
- Critère B: Les études exclues doivent être réunies dans un tableau, une liste ou une figure qui sera intégrée à l'article ou à un supplément.

- Critère C : Les raisons de l'exclusion des études sérieusement prises en considération doivent être exposées de manière suffisamment claire.
- Critère D : Le lecteur peut retracer aisément les études incluses et exclues dans la bibliographie, les références ou le supplément de l'article.
- Conditions d'attribution du score : 4 critères → 4 points, 3 → 3, 2 → 2, 1 → 1

EXEMPLE 15.4 (suite)

Dans la section « Résultats » de l'article, la première figure illustre le processus de sélection des études recensées, incluses et exclues, de même que les motifs d'exclusion. Une énumération des études incluses et un résumé des caractéristiques de ces études sont présentés dans un premier tableau. Cependant, les 22 études exclues n'apparaissent pas dans la liste de références.

Les critères A, B et C sont respectés, et le critère D est partiellement respecté ; le score est de 3 points sur 4.

6. Les caractéristiques des études incluses sont-elles indiquées ?

Les données portant sur les sujets qui ont participé aux études originales, les interventions qu'ils ont reçues et les résultats doivent être regroupés, sous forme de tableau, par exemple. La distribution des données sur les caractéristiques des sujets de toutes les études analysées doit y figurer.

- Critère A : Les données portant sur les sujets qui ont participé aux études originales, les interventions qu'ils ont reçues et les résultats sont regroupés, sous forme de tableau, par exemple.
- Critère B : Les auteurs précisent la distribution des données sur les caractéristiques pertinentes des sujets des études analysées (par exemple, l'âge, l'origine ethnique, le sexe, les données socioéconomiques pertinentes, la nature, la durée et la gravité de la maladie, d'autres maladies).
- Critère C : L'information fournie semble complète et exacte.
- Conditions d'attribution du score : 3 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4 (suite)

Dans la section « Résultats » de l'article, on trouve dans un premier tableau un résumé des caractéristiques de chaque étude recensée telles que :

Source (auteur principal et année), durée de l'étude, nombre de participants, caractéristiques de base des participants (âge, proportion de femmes), durée du diabète, niveau de base du pourcentage d'hémoglobine glycosylée, description de la médication administrée au groupe traité et au groupe témoin (hypoglycémiant ou placebo).

Les résultats de chaque essai font partie des deux figures qui présentent la synthèse quantitative de la principale mesure de l'effet glycémique, soit la différence de variation de l'hémoglobine glycosylée entre les groupes traités et témoins.

Les 3 critères sont respectés ; le score est de 4 points sur 4.

7. La qualité scientifique des études incluses a-t-elle été évaluée et consignée?

Les méthodes d'évaluation déterminées *a priori* doivent être indiquées (par exemple, pour les études sur l'efficacité pratique, le choix de n'inclure que les essais cliniques randomisés à double insu avec placebo ou de n'inclure que les études où l'affectation des sujets aux groupes d'étude était dissimulée); pour d'autres types d'études, d'autres critères d'évaluation seront à prendre en considération.

- Critère A: Les méthodes *a priori* sont indiquées.
- Critère B: La qualité scientifique des études incluses semble bonne.
- Critère C: Le niveau de preuve est exposé, dûment reconnu ou pris en considération.
- Critère D: La qualité des preuves est évaluée ou classée en fonction d'outils d'évaluation de la preuve. Un outil d'évaluation de la preuve est un instrument qui sert à établir le niveau de preuve (par exemple, l'outil GRADE [*Grading of Recommendations Assessment, Development and Evaluation*]).
- Conditions d'attribution du score: 4 critères → 4 points, 3 → 3, 2 → 2, 1 ou 0 → 1

EXEMPLE 15.4 (suite)

Dans la section «Méthode» de l'article, on nous renseigne sur les méthodes d'évaluation déterminées *a priori* pour être incluses :

Doit rapporter des données originales tirées d'essais cliniques hasardisés, réalisés dans une population d'adultes (femmes non enceintes) atteints de diabète de type 2, chez laquelle l'effet sur le niveau d'hémoglobine glycosylée (HbA_{1c}) a été comparé entre des sujets traités par un mimétique de l'incrétine et des sujets traités par un hypoglycémiant ou recevant un placebo. Les deux réviseurs ont exclu les essais d'une durée inférieure à 12 semaines.

Se reportant à la section «Méthode», on peut lire la description d'une méthode systématique d'évaluation de la validité des études. Il s'agit de différents critères qui se présentent comme suit :

Les différences quant aux caractéristiques de base entre les groupes, la description du masquage de l'allocation, la modalité d'analyse des données (en intention de traitement, tous les patients traités, analyse des participants ayant des données complètes) et la proportion de sujets perdus au suivi ont servi à déterminer la qualité des études.

Les critères A, B et C sont respectés ; le score est de 3 points sur 4.

8. La qualité scientifique des études incluses dans la revue a-t-elle été utilisée adéquatement dans la formulation des conclusions?

Les résultats de l'évaluation de la rigueur méthodologique et de la qualité scientifique des études incluses doivent être pris en considération dans l'analyse et les conclusions de la revue, et formulés explicitement dans les recommandations.

- Critère A: Les auteurs ont tenu compte de la qualité scientifique dans l'analyse et les conclusions de la revue.

- Critère B: La qualité scientifique est formulée explicitement dans les recommandations.
- Critère C: Les conclusions sont orientées vers la production de guides de pratique.
- Critère D: L'énoncé de consensus clinique laisse entrevoir la révision ou la confirmation des recommandations de pratique.
- Conditions d'attribution du score: 4 critères → 4 points, 3 → 3, 2 → 2, 1 ou 0 → 1

EXEMPLE 15.4 (suite)

La lecture de la discussion a permis à Lisa de rapporter au groupe comment la qualité scientifique des études incluses a été prise en compte dans la formulation des conclusions :

La discussion des inférences cliniques présente des arguments basés sur la validité interne, soit: la connaissance de l'efficacité et de l'innocuité à long terme est limitée par la courte durée de la plupart des essais; l'efficacité sur la glycémie a pu être surestimée, car peu d'essais ont réalisé une vraie analyse en intention de traitement, et les proportions des sujets perdus au suivi étaient élevées. La discussion comporte également des arguments basés sur la validité externe: le faible niveau de base de l'hémoglobine glycosylée des sujets peut contribuer à la faible ampleur de l'effet; les effets différentiels selon l'ethnie sont inconnus; les résultats ne s'appliquent qu'aux adultes. Les résultats sont interprétés à la lumière de nombreuses évidences: le rôle du faible niveau de base de l'hémoglobine glycosylée des sujets; l'effet intéressant sur la glycémie postprandiale; la comparaison aux traitements habituels; l'effet sur le poids et son lien possible avec les nausées; les épisodes d'hypoglycémie peu fréquents; l'effet de la combinaison avec d'autres hypoglycémifiants; les effets indésirables; l'impact possible d'une augmentation des infections urinaires dans une population de diabétiques plus vulnérables aux infections.

Seul le critère A est respecté; le score est de 1 point sur 4.

9. Les méthodes utilisées pour combiner les résultats des études sont-elles appropriées?

Si l'on veut regrouper les résultats des études, il faut effectuer un test d'homogénéité afin de s'assurer qu'elles sont combinables (par exemple, khi carré ou I^2). S'il y a hétérogénéité, il faut utiliser un modèle d'effets aléatoires et (ou) vérifier si la nature des données cliniques justifie la combinaison (la combinaison est-elle raisonnable?).

- Critère A: Les auteurs exposent les critères à partir desquels ils ont déterminé que les études analysées étaient assez semblables pour être combinées.
- Critère B: Dans le cas des résultats regroupés, les auteurs ont fait un test d'homogénéité pour s'assurer que les études étaient combinables.
- Critère C: Les auteurs ont pris acte du caractère hétérogène (ou non) des études.
- Critère D: S'il y a hétérogénéité, les auteurs ont utilisé un modèle d'effets aléatoires et (ou) vérifié si la nature des données justifiait la combinaison.

- Critère E: S'il y a homogénéité, les auteurs exposent la justification ou le test statistique.
- Conditions d'attribution du score: 4 ou 5 critères → 4 points, 3 → 3, 2 → 2, 1 ou 0 → 1

EXEMPLE 15.4 (suite)

La lecture des méthodes a permis à Lisa de prendre connaissance des éléments de la synthèse quantitative. Elle les résume ainsi au groupe :

La principale mesure de l'effet glycémique est la différence de variation de l'hémoglobine glycosylée entre les groupes traités et témoins; d'autres mesures d'effet évaluent l'innocuité de l'intervention. Pour combiner les résultats à partir de variables continues comme l'hémoglobine glycosylée, une moyenne pondérée des différences de variation entre les groupes traités et témoins est calculée avec son intervalle de confiance à 95%. Pour les variables dichotomiques, comme la proportion des sujets ayant présenté de l'hypoglycémie, les rapports de proportions sont calculés avec leur intervalle de confiance à 95%. Les synthèses quantitatives ont été calculées à partir du modèle à effets aléatoires en pondérant les résultats selon l'inverse de la variance et la variation interétudes, en appliquant la méthode de DerSimonian et Laird. La statistique I^2 décrit la proportion de la variation totale entre les essais, attribuable à l'hétérogénéité¹⁷. Afin d'explorer l'hétérogénéité, des analyses de sensibilité et de sous-groupes basées sur le type de traitement (agonistes du GLP-1 et inhibiteurs de la DPP4), le type de groupe témoin (ayant reçu un hypoglycémiant ou un placebo) et la durée de l'intervention (12 semaines *versus* > 12 semaines) ont été réalisées. Pour certaines mesures d'effets (glycémie postprandiale, profil lipidique), aucune synthèse quantitative n'est faite en raison de la diversité des méthodes utilisées ou de l'insuffisance des données.

Les critères B et D sont respectés; le score est de 2 points sur 4.

10. La probabilité d'un biais de publication a-t-elle été évaluée?

Une évaluation du biais de publication doit comprendre une association d'outils graphiques (diagramme de dispersion des études ou autre test) et (ou) des tests statistiques (par exemple, le test de régression d'Egger).

- Critère A: Prise en compte du biais de publication
- Critère B: Outils graphiques (par exemple, un diagramme de dispersion des études)
- Critère C: Tests statistiques (par exemple, le test de régression d'Egger)
- Conditions d'attribution du score: 3 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4 (suite)

Marie-Louise rapporte qu'il n'y a aucune évaluation du biais de publication dans l'article; le score est de 1 point sur 4.

11. Les conflits d'intérêts ont-ils été déclarés?

Les sources possibles de soutien doivent être déclarées, tant pour la revue systématique que pour les études qui y sont incluses.

- Critère A: Présentation des sources de soutien
- Critère B: Absence de conflit d'intérêts – On est ici dans la subjectivité; peut-être faudra-t-il y aller par déduction ou fouiller quelque peu.
- Critère C: Prise en compte ou énoncé des sources de soutien ou des conflits d'intérêts dans les principales études incluses
- Conditions d'attribution du score: 3 critères → 4 points, 2 → 3, 1 → 2, 0 → 1

EXEMPLE 15.4 (suite)

Après avoir lu l'article, Lisa rapporte que les auteurs y indiquent que toutes les études incluses ont été financées par des compagnies pharmaceutiques et que le rôle du commanditaire a été clairement divulgué dans seulement deux études.

À la fin de l'article, les auteurs font état d'une absence de conflit d'intérêts, mentionnent leurs sources de soutien et indiquent qu'elles n'ont joué aucun rôle à aucune étape de la revue.

Les trois critères sont respectés; le score est de 4 points sur 4.

Bilan

Pour les 11 critères de la grille, le score maximal est de 44. Lisa obtient suite à son analyse un score de 31 sur 44. Ce score fournit un indice de la qualité globale d'une synthèse systématique.

15.5 La modification de sa pratique à la suite de la lecture d'un article de synthèse systématique

Lisa et son équipe, tout comme la majorité des cliniciens, choisissent des articles pertinents dans le contexte de leur pratique. Parfois, leurs lectures entraîneront un changement dans leurs habitudes. Après avoir porté un jugement sur la force de la preuve scientifique, certaines considérations devraient être prises en compte lorsqu'un clinicien évalue la nécessité de modifier sa pratique en fonction des conclusions d'un article de synthèse systématique.

D'abord, les patients de sa pratique sont-ils comparables aux patients — ou à certains sous-groupes de patients — des essais ayant fait l'objet de la synthèse systématique? Les contraintes liées à sa pratique rendent-elles possible une modification de celle-ci? En d'autres mots, l'intervention est-elle abordable et accessible, l'expertise nécessaire est-elle disponible dans son milieu de pratique? Les bénéfices escomptés et les risques d'effets indésirables sont-ils connus pour des sous-groupes de patients dont

les caractéristiques correspondent à ceux de sa pratique? Enfin, il lui faudra en outre tenir compte des valeurs et des préférences de ses patients au moment de décider d'une option thérapeutique.

Après avoir discuté de l'article de synthèse systématique, Lisa et ses collègues doivent décider de la nécessité de modifier leur pratique en fonction des résultats de cette méta-analyse. Ils tirent cette conclusion : les mimétiques de l'incrétine s'avèrent une option thérapeutique pour certains patients adultes atteints de diabète de type 2, en particulier ceux dont le diabète est léger, qui tolèrent mal les traitements habituels, et qui ne sont pas sujets aux infections ni aux céphalées; cependant, le suivi de ces patients appelle à la vigilance.

15.6 La Collaboration Cochrane

Déjà en 1972, celui qui a donné son nom à la Collaboration Cochrane, l'épidémiologiste britannique Archie Cochrane, soulignait le manque flagrant de preuves scientifiques dans le domaine des soins de santé. Après la réalisation d'un projet précurseur visant la constitution d'un registre de tous les essais cliniques en médecine périnatale, la Collaboration Cochrane est officiellement fondée en 1993¹⁸.

La Collaboration Cochrane a pour mission d'aider à la prise de décisions dans le domaine de la santé. Elle regroupe des professionnels de la santé, des cliniciens, des chercheurs et des consommateurs et compte environ 50 groupes de travail répartis sur les cinq continents, tous à la base des productions de la Collaboration Cochrane. Elle produit et distribue des synthèses systématiques sur les interventions en santé et fait la promotion de la recherche de preuves scientifiques, principalement sous forme d'essais cliniques. Préparées par les membres des groupes de travail, sous la supervision d'équipes éditoriales, les synthèses systématiques Cochrane doivent leur renommée à l'application de rigoureux standards de qualité.

On trouve les productions de la Collaboration Cochrane dans une base de données appelée Cochrane Library¹⁹. Celle-ci contient plusieurs banques de données, dont deux spécialisées en médecine basée sur les faits (*evidence-based medicine*), mises à jour quatre fois par année. La Cochrane Database of Systematic Reviews contient le texte intégral de synthèses systématiques portant sur l'efficacité des soins de santé, alors que le Cochrane Central Register of Controlled Trials recense les références bibliographiques compilées d'essais cliniques.

En résumé

Les articles de synthèse systématique sont des textes scientifiques dans lesquels les auteurs résumement et analysent les études originales traitant d'un sujet donné en suivant un protocole méthodologique rigoureux. Les synthèses systématiques peuvent comporter une synthèse quantitative des résultats de plusieurs études; on parle alors de méta-analyse. Les articles de synthèse systématique permettent de dresser le bilan des connaissances sur une question donnée, d'établir des guides de pratique et de suggérer des pistes de recherche.

Les synthèses systématiques se caractérisent par les aspects suivants: l'utilisation d'une stratégie de recherche visant à recenser toutes les études pertinentes sur un sujet; l'emploi de critères explicites servant à justifier l'inclusion et l'exclusion des études ainsi qu'à évaluer la qualité des études retenues; l'exploration des sources possibles d'hétérogénéité parmi les résultats d'études; la synthèse quantitative de

l'estimation de l'effet lorsque cela est pertinent (dans le cas des méta-analyses); et la présentation de la méthodologie utilisée dans l'article de synthèse systématique.

Les outils AMSTAR et R-AMSTAR proposent une grille d'évaluation de la qualité méthodologique des synthèses systématiques et une liste de critères d'attribution d'un score. Par la suite, le lecteur doit évaluer les conclusions de l'article de synthèse systématique pour déterminer si elles sont applicables dans le contexte de sa pratique clinique; il pourra alors choisir de modifier cette dernière.

La Collaboration Cochrane propose, dans ses banques de données, une vaste collection de synthèses systématiques portant sur les interventions en santé. Les synthèses systématiques réalisées par les groupes Cochrane respectent des standards de rigueur méthodologique élevés.

Exercices



Voir la section complète des exercices sur le Web.

Vrai ou faux.

1. Un biais de publication peut altérer les résultats d'un article de synthèse systématique parce que les études dont les résultats sont statistiquement significatifs ont moins de chance d'être publiées.
2. Le test du khi carré d'hétérogénéité est utile pour évaluer la présence d'hétérogénéité alors que les analyses par sous-groupes, les analyses de sensibilité et la méta-régression sont employées pour chercher les causes d'hétérogénéité.
3. Les méta-analyses permettent d'améliorer la puissance statistique parce qu'en augmentant la taille des échantillons, elles accroissent la précision des résultats.
4. Dans une synthèse systématique, s'il s'avère impossible d'expliquer l'hétérogénéité entre les résultats des études, la validité des conclusions en sera limitée; il devient alors préférable de conclure que l'hétérogénéité est trop grande pour combiner les résultats.
5. Il n'est pas essentiel que les auteurs d'une synthèse systématique décrivent les critères qu'ils ont utilisés pour l'inclusion et l'exclusion des études qu'ils ont retenues.

Pour les réponses, voir le corrigé à la fin du manuel.

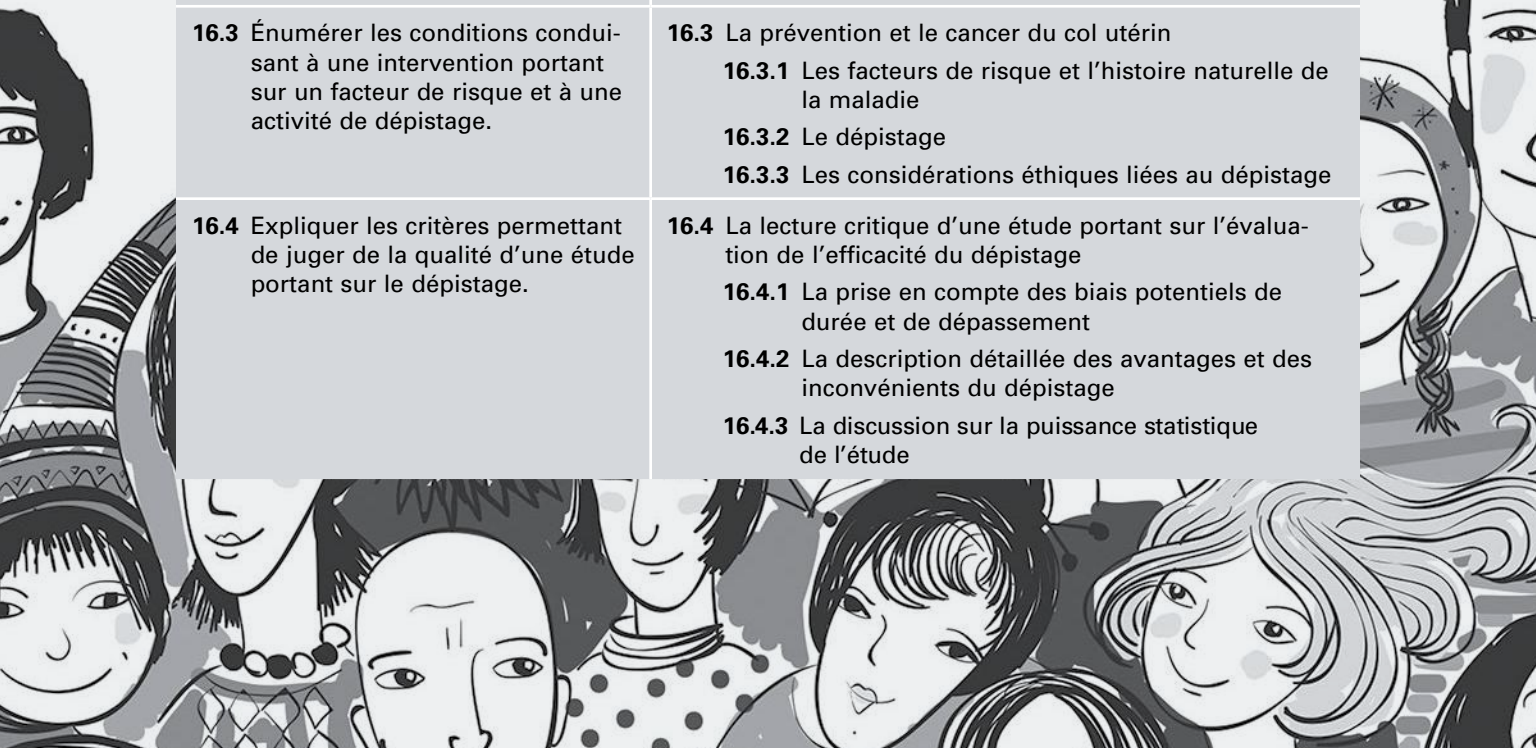
Collaboration à la 3^e édition :

Louise Moreault
Clément Beaucage
André Simpson

Au terme de ce chapitre, vous pourrez ;

- expliquer les concepts épidémiologiques qui permettent d'établir les bases du jugement clinique au regard des activités de prévention et en particulier du dépistage ;
- juger de l'utilité pour votre pratique des études portant sur l'efficacité du dépistage.

Objectifs spécifiques	Contenu
16.1 Définir le rôle du clinicien en prévention.	16.1 Le rôle du personnel soignant en prévention clinique
16.2 Expliquer les liens entre l'histoire naturelle d'une maladie, les niveaux de prévention et les catégories d'interventions préventives.	16.2 L'évolution des pratiques cliniques préventives 16.2.1 La méthode du U.S. Preventive Services Task Force (USPSTF) 16.2.2 La stratégie d'utilisation 16.2.3 Un retour sur le cas de Jeanne
16.3 Énumérer les conditions conduisant à une intervention portant sur un facteur de risque et à une activité de dépistage.	16.3 La prévention et le cancer du col utérin 16.3.1 Les facteurs de risque et l'histoire naturelle de la maladie 16.3.2 Le dépistage 16.3.3 Les considérations éthiques liées au dépistage
16.4 Expliquer les critères permettant de juger de la qualité d'une étude portant sur le dépistage.	16.4 La lecture critique d'une étude portant sur l'évaluation de l'efficacité du dépistage 16.4.1 La prise en compte des biais potentiels de durée et de dépassement 16.4.2 La description détaillée des avantages et des inconvénients du dépistage 16.4.3 La discussion sur la puissance statistique de l'étude





Mise en situation

Jeanne consulte Lisa, omnipraticienne à la polyclinique Milo, car, depuis quelques mois, elle se sent essoufflée à la suite d'efforts modérés. Elle a 56 ans, est mariée et a 3 enfants. Jeanne fume 20 cigarettes par jour depuis 35 ans; elle tousse et crache régulièrement depuis plus de 10 ans. Elle a été ménopausée à l'âge de 49 ans. Son dernier examen de dépistage du cancer du col utérin a été effectué il y a 20 ans, soit après l'accouchement de son dernier enfant. Les antécédents familiaux révèlent que son père est décédé d'un infarctus du myocarde à l'âge de 50 ans et que 2 de ses sœurs souffrent d'hypercholestérolémie.

L'examen physique, incluant l'examen gynécologique et des seins, est normal, mis à part quelques râles aux bases pulmonaires. Le pouls et la pression artérielle se situent dans les limites normales. Lisa explique à Jeanne que son essoufflement est probablement dû à une maladie pulmonaire obstructive chronique liée à sa consommation de tabac.

Lisa entreprend une investigation et un traitement appropriés à l'état de santé respiratoire

de sa patiente. Elle envisage de lui recommander certaines mesures préventives.

Pertinence clinique

Le rôle premier de Lisa est de répondre à la demande de Jeanne. Cependant, dans un contexte de pratique, Lisa sera aussi appelée à offrir des services préventifs. Plusieurs ignorent que les professionnels de la santé consacrent une partie de leur temps à des activités d'éducation des patients, de dépistage et de suivi de personnes présentant des facteurs de risque. Ainsi, le médecin peut aisément allouer de 15 à 25% de son temps à ces activités. On estime que 79% des Canadiens voient un médecin au moins une fois par année¹. On peut imaginer que la situation est sensiblement la même pour d'autres intervenants tels que les infirmières et les pharmaciens. Cette fréquence des contacts constitue un terrain fertile pour établir une intervention de nature préventive. Lisa et ses collègues occupent en fait une position stratégique dans la prévention des maladies et leur détection précoce.

16.1 Le rôle du personnel soignant en prévention clinique

Le concept même de santé est en profonde mutation; de nouveaux problèmes, tels l'augmentation du nombre de cas d'obésité et le vieillissement de la population, ne permettent plus d'aborder les soins de santé de la même façon qu'auparavant. En

outre, la société investit beaucoup de ressources dans le système de soins et s'attend à ce que la santé de la population s'en trouve améliorée. Les gens veulent que leur qualité de vie soit maintenue; de nombreux patients demandent aux praticiens de leur dire comment prévenir l'apparition de la maladie ou de leur recommander le traitement le plus efficace au meilleur prix possible. Le rôle du personnel soignant doit évoluer et s'adapter à ces nouvelles réalités.

En établissant une relation de confiance avec son patient, le professionnel de la santé peut s'assurer de la continuité des interventions, pierre angulaire des soins de première ligne et, du coup, donner beaucoup de crédibilité aux mesures préventives.

16.2 L'évolution des pratiques cliniques préventives

Pendant longtemps, à l'occasion d'une consultation comme celle de Jeanne auprès de Lisa, le médecin aurait prescrit une série d'examens pour dresser un bilan de santé annuel. Encore aujourd'hui, certains patients nostalgiques peuvent demander à leur médecin de leur faire un bon *check-up*. La prescription du bilan de santé annuel n'est cependant plus indiquée pour les raisons suivantes.

- La nature et la fréquence des examens ne tenaient pas compte des besoins des différents groupes d'âge.
- Les preuves que les examens ou les activités de dépistage effectués à l'occasion d'un bilan de santé constituaient des mesures valables et efficaces étaient insuffisantes.
- Ces activités de dépistage étaient répétées annuellement, alors que nombre d'entre elles n'exigeaient aucunement une telle fréquence.

Beaucoup d'examens à visée préventive ne sont pas explicitement demandés par Jeanne lors de sa visite à la polyclinique. De ce fait, Lisa ne peut les lui suggérer que si leur pertinence repose sur des preuves scientifiques et sur la démonstration que Jeanne en retirera plus d'avantages que d'inconvénients.

C'est cette préoccupation qui a conduit à la mise sur pied, en 1976, du Groupe d'étude canadien sur les soins de santé préventifs (GECSSP)². Dans la foulée, d'autres pays ont créé des groupes similaires. Notamment, en 1984, on a formé aux États-Unis le U.S. Preventive Services Task Force (USPSTF). Ces groupes tentent de préciser l'efficacité et la pertinence de certains actes préventifs en révisant la qualité des preuves scientifiques afin que les professionnels de la santé disposent de lignes directrices claires en matière de services préventifs. Les deux groupes ont acquis une réputation d'envergure mondiale grâce à la rigueur de leurs analyses. Ils reconnaissent que la performance des actes préventifs doit être évaluée systématiquement et de manière objective. Ainsi, leur principale recommandation est d'abandonner le bilan de santé annuel au profit d'une démarche sélective basée sur l'âge et le sexe de chaque personne.

Le GECSSP³ et le USPSTF⁴ utilisent des méthodes comparables. Ces groupes ont émis des avis sur plusieurs dizaines de problèmes de santé qui peuvent présenter quelques différences, bien que, dans l'ensemble, ils soient le plus souvent convergents. Ces avis se retrouvent sur leur site Internet respectif⁵. Depuis quelques années,

en Amérique du Nord, le USPSTF est le groupe le plus actif. Pour cette raison, sa méthode et ses recommandations seront présentées ici.

16.2.1 La méthode du U.S. Preventive Services Task Force (USPSTF)

Tout comme dans le cas du GECSSP, les recommandations du USPSTF visent des interventions préventives centrées sur l'identification des facteurs de risque (prévention primaire) et sur le **dépistage** (prévention secondaire). Le USPSTF établit ses recommandations sur la base de preuves scientifiques existantes. Deux aspects importants des problèmes de santé retenus sont traités :

- le fardeau de souffrance engendré par le problème de santé pour lequel une intervention est proposée ;
- l'évaluation de l'efficacité de l'intervention préventive proposée.

L'évaluation du fardeau de souffrance

Le fardeau de souffrance associé à chaque problème de santé est établi en considérant deux facteurs : l'impact sur l'individu au chapitre des années de vie perdues, de l'inconfort, de l'incapacité et des coûts entraînés par la thérapeutique ainsi que des effets pour sa famille, de même que les répercussions sur la société, ces deux facteurs étant évalués à partir des données de mortalité, de morbidité et de coûts afférents.

L'évaluation de l'efficacité de l'intervention préventive

L'efficacité de l'intervention préventive est évaluée en fonction de la qualité des preuves scientifiques que constituent les résultats d'études recueillies dans la littérature scientifique. Cette évaluation est effectuée en tenant compte de trois aspects :

- les risques et les avantages que comporte l'intervention ;
- la sensibilité, la spécificité et la valeur prédictive de l'examen utilisé, le cas échéant ;
- la sécurité, la simplicité et le coût de l'intervention, ainsi que son acceptabilité par le patient.

Le USPSTF propose de classifier les preuves scientifiques existantes en trois catégories, selon le niveau de certitude relatif au bénéfice réel de l'intervention préventive. Le tableau 16.1 présente cette classification.

TABLEAU 16.1 ■ Les niveaux de certitude concernant le bénéfice réel

Niveau de certitude*	Description
Élevé	Les preuves scientifiques existantes incluent habituellement des résultats solides provenant d'études bien construites, bien menées dans des populations représentatives des soins de santé offerts en première ligne. Ces études évaluent les effets de l'intervention préventive sur la santé. De plus, cette conclusion n'est pas susceptible d'être significativement modifiée par les résultats d'études ultérieures. →

TABLEAU 16.1 ■ Les niveaux de certitude concernant le bénéfice réel (suite)

Niveau de certitude*	Description
Modéré	<p>Les preuves scientifiques existantes sont suffisantes pour déterminer les effets de l'intervention préventive sur la santé, mais le niveau de confiance de l'estimation est contraint par des facteurs, tels que :</p> <ul style="list-style-type: none"> • le nombre, la taille ou la qualité des études ; • l'incohérence des résultats entre les études ; • les limites dans la capacité de généraliser les résultats au contexte de soins de première ligne ; • l'incohérence dans la chaîne d'évidence. <p>À mesure que plus d'information deviendra disponible, l'ampleur et le sens des effets observés pourraient changer, et ces changements pourraient s'avérer suffisamment importants pour en modifier la conclusion.</p>
Faible	<p>Les preuves scientifiques existantes sont insuffisantes pour évaluer les effets sur la santé :</p> <ul style="list-style-type: none"> • le nombre ou la taille limités des études ; • les fautes importantes dans le devis ou dans les méthodes utilisées dans les études ; • l'incohérence des résultats entre les études ; • les lacunes dans la chaîne d'évidence ; • les résultats non généralisables au contexte de soins de première ligne ; • le manque d'informations sur des résultats importants pour la santé. <p>Des informations supplémentaires permettraient d'estimer les effets sur la santé.</p>

* Le USPSTF définit le niveau de certitude comme étant «la probabilité que l'évaluation du USPSTF quant au bénéfice réel soit exacte». Le bénéfice réel est défini comme étant le bénéfice attendu moins les risques de l'intervention préventive lorsqu'elle est implantée dans le contexte des soins de première ligne. Le USPSTF attribue un niveau de certitude s'appuyant sur l'ensemble des preuves scientifiques disponibles pour évaluer le bénéfice réel d'une intervention préventive.

Source : Traduit et adapté de U.S. Preventive Services Task Force (USPSTF) (2007). *Update on methods: Estimating certainly and magnitude of net benefit*, [en ligne], www.uspreventiveservicestaskforce.org/Page/Name/update-on-methods-estimating-certainty-and-magnitude-of-net-benefit (page consultée le 20 décembre 2016).

Les recommandations

Chaque recommandation du USPSTF concerne un type d'intervention préventive pour un problème de santé particulier et vise un ou plusieurs groupes d'âge précis. La recommandation indique si l'intervention devrait être incluse à l'occasion d'une visite médicale ou offerte systématiquement aux patients. Les recommandations sont une représentation utile pour le clinicien du niveau de certitude et de l'ampleur du bénéfice réel de la mesure préventive⁶. Depuis 2012, elles sont regroupées en quatre catégories suivant l'ordre alphabétique de A à D, comme le montre le tableau 16.2. De plus, une catégorie I a été ajoutée pour les cas où les preuves s'avèrent insuffisantes.

Ainsi, les recommandations de catégories A et D s'appuient généralement sur une preuve de niveau de certitude élevé pour affirmer qu'il y a ou non un bénéfice réel. Des interventions font l'objet de recommandations de catégorie B qui semblent plus

ou moins affirmatives selon le niveau des preuves qui les appuient; de telles situations peuvent se produire lorsque le USPSTF constate certaines lacunes quant à la force des preuves scientifiques existantes ou quant à l'efficacité de l'intervention préventive elle-même dans un contexte de soins de première ligne.

La recommandation de catégorie C peut sembler peu utile, car elle contribue moins clairement à la prise de décision. Toutefois, le USPSTF y a recours pour marquer sa prudence, par exemple lorsque le changement de pratique est important et qu'il doit être appuyé par de fortes preuves. Il peut aussi y recourir pour éviter un effet d'étiquetage (soit le fait d'identifier des personnes comme anormales) ou encore pour empêcher que des examens coûteux et d'utilité inconnue fassent gonfler la facture des services de santé. Une recommandation de catégorie C peut également servir à attirer l'attention sur des problèmes de santé qui entraînent un fardeau de souffrance considérable et sur des problèmes particuliers touchant des groupes à risque.

TABLEAU 16.2 ■ La classification des recommandations émises par le USPSTF

Grade	Définition	Suggestion pour la pratique
A	Le USPSTF recommande l'intervention. Il y a un niveau de certitude élevé que le bénéfice réel soit substantiel.	Offrir ou fournir cette intervention préventive.
B	Le USPSTF recommande cette intervention préventive. Il y a un niveau de certitude élevé que le bénéfice réel soit modéré ou il y a un niveau de certitude modéré que le bénéfice réel soit de modéré à substantiel.	Offrir ou fournir cette intervention préventive.
C	Le USPSTF recommande d'offrir ou de fournir cette intervention préventive de manière sélective aux patients en s'appuyant sur le jugement du professionnel et sur les préférences du patient. Il existe au moins un niveau modéré de certitude qu'il y ait un faible bénéfice.	Offrir ou fournir cette intervention préventive à des patients choisis en fonction de situations ou de caractéristiques personnelles.
D	Le USPSTF recommande de ne pas offrir cette intervention préventive. Il y a un niveau de certitude modéré ou élevé que l'intervention n'ait pas de bénéfice réel ou que les risques surpassent les avantages.	Décourager l'utilisation de cette intervention préventive.
I	Le USPSTF conclut que les preuves scientifiques existantes sont actuellement insuffisantes pour évaluer le rapport entre les avantages et les risques liés à cette intervention préventive. Il y a absence d'évidence scientifique ou celle-ci est de pauvre qualité ou même contradictoire. Ainsi, le rapport entre les avantages et les risques ne peut être déterminé.	Lire la section sur les considérations cliniques du <i>USPSTF Recommendation Statement</i> . Si l'intervention préventive est offerte, les patients doivent comprendre le niveau d'incertitude dans le rapport entre les avantages et les risques.

Source: Traduit et adapté de U.S. Preventive Services Task Force (USPSTF) (2013). *Grade definitions*, [en ligne], www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions (page consultée le 23 juin 2016).

Le cas échéant, le USPSTF mentionne dans sa recommandation la fréquence optimale de l'examen ou de l'intervention préventive, l'existence de groupes à risque élevé qui pourraient tirer avantage d'examens plus fréquents et le moment où ces examens devraient être pratiqués. Les recommandations font l'objet d'un processus de révision systématique tous les cinq ans et de mises à jour au besoin.

16.2.2 La stratégie d'utilisation

La stratégie du USPSTF consiste à recourir à un ensemble de mesures de protection sanitaires. Cet ensemble est constitué d'interventions qui s'appliquent dans le cadre d'une visite médicale à certaines catégories d'âge ou à des groupes de patients à risque élevé. La périodicité de chacune des mesures préventives faisant partie d'un ensemble de mesures de protection sanitaires est liée à la connaissance plus ou moins étendue de l'histoire naturelle de la maladie et à certaines contraintes de coûts et de ressources.

Le USPSTF a élaboré une classification correspondant assez fidèlement aux activités cliniques habituelles d'un médecin. Cette classification regroupe les différents types d'interventions préventives en quatre catégories : les interventions de conseil, de dépistage, de vaccination et de chimioprophylaxie. Le tableau 16.3 donne des exemples de chacun des types d'intervention.

TABLEAU 16.3 ■ Des exemples d'interventions préventives selon leur catégorie

Catégorie d'intervention préventive	Exemples
Interventions de conseil	<ul style="list-style-type: none"> • Conseils pour l'arrêt tabagique • Conseils sur la pratique d'activité physique
Interventions de dépistage	<ul style="list-style-type: none"> • Recherche de sang dans les selles pour le dépistage du cancer colorectal • Utilisation de la mammographie pour le dépistage du cancer du sein
Interventions de vaccination	<ul style="list-style-type: none"> • Vaccination en prévention de l'hépatite B • Vaccination contre la fièvre jaune pour certains groupes de voyageurs
Interventions de chimioprophylaxie	<ul style="list-style-type: none"> • Prise d'acide folique chez la femme en âge de procréer en prévention des anomalies du tube neural chez le bébé • Prise d'antimalariques en prévention du paludisme

Les interventions de conseil, de vaccination et de chimioprophylaxie sont des interventions de prévention primaire, puisqu'elles visent à modifier certains facteurs de risque avant qu'un problème de santé ne se développe. Les interventions de dépistage sont des activités de prévention secondaire qui visent à détecter une maladie à un stade précoce et asymptomatique de son évolution pour en améliorer le pronostic.

Plusieurs interventions préventives suggérées par le USPSTF ou le GECSSP ont été largement adoptées par l'ensemble des professionnels de la santé. C'est le cas, par exemple, du dépistage du cancer du col utérin. Cependant, il existe souvent un fossé entre les recommandations d'activités préventives et la pratique clinique quotidienne. Afin d'améliorer la santé de la population, il importe donc que les praticiens mesurent mieux la portée et l'importance de traduire les lignes directrices de ces groupes en des gestes concrets pendant les rencontres avec les patients. Ainsi, Lisa est déjà au fait des travaux du USPSTF, et la visite de Jeanne la place devant la nécessité de faire plusieurs choix.

16.2.3 Un retour sur le cas de Jeanne

Dans une perspective globale de soins, Lisa se posera les questions suivantes quant à la pertinence d'offrir à Jeanne des interventions préventives en ce qui concerne certains problèmes de santé particuliers.

- Dans le cadre de la prévention du cancer du poumon, est-il pertinent de lui recommander de cesser de fumer?
- Dans le cadre de la prévention de la maladie coronarienne, devrait-on procéder chez elle au dépistage de l'hypercholestérolémie et lui recommander de cesser de fumer?
- Dans le cadre de la prévention du cancer du sein, doit-on lui recommander de passer une mammographie?
- Dans le cadre de la prévention du cancer du col utérin, est-il pertinent de lui recommander de subir une cytologie de dépistage (test de Papanicolaou)?

Le cas de Jeanne soulève donc des interrogations sur l'efficacité des mesures préventives. En effet, Lisa désire suggérer à sa patiente des interventions qui seront véritablement bénéfiques pour sa santé. Prenons pour exemple la prévention du cancer du col utérin.

16.3 La prévention et le cancer du col utérin

16.3.1 Les facteurs de risque et l'histoire naturelle de la maladie

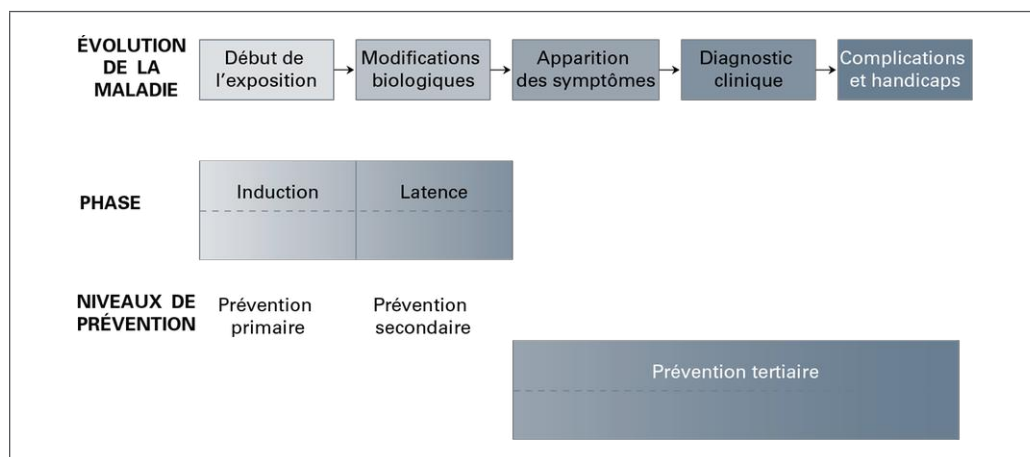
Le cas de Jeanne suscite tout naturellement une discussion sur le cancer du col utérin. L'étude de ce problème de santé a permis de déceler certains facteurs de risque de la maladie, c'est-à-dire des conditions qui sont associées à une probabilité accrue d'en être atteinte. Ces facteurs de risque sont la précocité des relations sexuelles, la multiplicité des partenaires sexuels, le tabagisme et les antécédents d'infection transmissible sexuellement et par le sang, en particulier d'infection par le virus du papillome humain (VPH). Le cancer du col utérin est le premier type de cancer à avoir été associé au VPH; ce virus est présent dans plus de 99 % des cas, ce qui en fait le principal facteur de risque⁷.

L'étude du cancer du col utérin a permis d'en décrire assez précisément l'histoire naturelle, soit les différentes phases qui précèdent le développement d'une lésion néoplasique invasive du col de l'utérus. Ainsi, à ce site, des modifications cellulaires

appelées « dysplasies » peuvent apparaître précocement pour engendrer, après environ 10 ans, un cancer localisé appelé « carcinome *in situ* ». En l'absence d'intervention, ce dernier progresse souvent vers un cancer envahissant⁸ pendant la décennie suivante.

Lisa doit connaître les facteurs de risque et les phases de développement de la maladie pour pouvoir effectuer toute démarche de prévention. Cette prévention peut être primaire, secondaire ou tertiaire. La **prévention primaire** consiste à reconnaître les principaux facteurs de risque auxquels l'individu est exposé et à lui offrir des conseils appropriés si ces facteurs sont modifiables ou encore à lui proposer la vaccination si cela s'avère indiqué. La détection de maladies à un stade où le patient est encore asymptomatique est appelée « dépistage »; il s'agit de la **prévention secondaire**. Enfin, la **prévention tertiaire** consiste à intervenir auprès du patient qui présente des symptômes afin de freiner l'évolution de la maladie. La figure 16.1 représente l'histoire naturelle de la maladie et les trois niveaux de prévention possibles selon les étapes de son évolution et les phases de son développement, soit la **phase d'induction** et la phase de latence.

FIGURE 16.1 ■ L'histoire naturelle de la maladie et les niveaux de prévention



Les questions auxquelles Jeanne répond dans le bureau de Lisa visent à déceler les facteurs de risque qui lui sont propres, notamment le tabagisme. Une intervention effectuée dans le but de l'amener à cesser de fumer s'inscrit dans une approche de prévention primaire du cancer du col utérin. Par contre, la vaccination contre le VPH (offerte aux jeunes femmes de 9 à 26 ans) n'est actuellement pas indiquée pour Jeanne, car il n'existe encore aucune preuve de son efficacité pour les femmes de son âge⁹. La recommandation de Lisa d'effectuer une cytologie du col utérin constitue une activité de dépistage; il s'agit là de prévention secondaire. Évidemment, pour qu'il y ait dépistage, le patient doit se trouver dans une période asymptomatique, dite **phase de latence**. Cette période doit être suffisamment longue pour permettre de détecter une modification biologique avant que la maladie ne se manifeste. Dans le cas du cancer du col utérin, la phase de latence correspond aux stades de dysplasie et de carcinome *in situ*, au cours desquels les lésions demeurent au stade cellulaire, sans envahissement ni symptômes.

16.3.2 Le dépistage

Une définition

Le dépistage consiste donc à reconnaître une maladie à un stade précoce asymptomatique à l'aide d'une activité de dépistage ou d'un examen déterminé. Ainsi, les activités de dépistage permettent de repérer, dans un groupe de gens asymptomatiques, ceux qui peuvent être porteurs d'une maladie. L'axiome suivant constitue la base rationnelle du dépistage : le traitement a plus de chances de se révéler efficace s'il est entrepris à un stade précoce de la maladie, alors que le processus pathologique peut être réversible et que des complications ne sont pas encore survenues. De la même façon, il est logique de penser qu'une thérapie instaurée dans une phase asymptomatique de la maladie sera probablement plus simple et plus facilement applicable que s'il s'agit de traiter un cas clinique avancé. Il importe de mentionner que l'efficacité d'un dépistage (c'est-à-dire sa capacité de diminuer le taux de mortalité attribuable à la maladie recherchée) doit avoir été démontrée par des essais cliniques hasardisés (ECH) de bonne qualité.

Il faut aussi préciser que les examens ou les activités de dépistage ne permettent pas de poser un diagnostic final. Les individus identifiés comme positifs au dépistage doivent subir des examens complémentaires (diagnostiques) plus spécifiques avant que ne soit établie la présence de la maladie. Ainsi, lorsqu'une cytologie du col utérin indique la présence de dysplasies ou de cellules anormales, une colposcopie* avec biopsie de la zone suspecte permettra de confirmer ou d'infirmer la présence de la maladie. Le cas échéant, il existe des traitements reconnus efficaces qui pourront être administrés.

Les caractéristiques des maladies dépistables

En ce qui concerne le dépistage du cancer du col utérin, le USPSTF s'est penché sur deux interventions : la cytologie cervico-vaginale et la recherche du VPH. Dans le présent chapitre, la recommandation d'une cytologie du col utérin tous les trois ans chez les femmes âgées de 21 à 65 ans sera retenue¹⁰. Cet examen demeure recommandé malgré la vaccination contre le VPH, puisque celle-ci ne permet pas d'éradiquer la maladie¹¹.

Avant d'effectuer une activité de dépistage comme la cytologie cervico-vaginale, certaines conditions doivent être remplies. En effet, il ne suffit pas de bien connaître l'histoire naturelle de la maladie et les facteurs de risque ; on doit aussi prendre en considération le fardeau de souffrance lié au problème de santé à dépister. De façon générale, les caractéristiques particulières de la maladie doivent justifier l'intervention de prévention secondaire. Pour mesurer ce fardeau de souffrance, il faut d'abord connaître la fréquence de la maladie dans la population.

Lisa sait que le cancer du col utérin occupe, dans le monde¹², le quatrième rang pour ce qui est des néoplasies chez les femmes. Par ailleurs, il importe également de

* La colposcopie est un examen du col utérin pratiqué à l'aide d'un appareil optique placé dans le vagin afin de voir le col à plus fort grossissement.

connaître les conséquences graves qu'entraîne la maladie, c'est-à-dire la morbidité et la mortalité qui y sont rattachées. Ainsi, Lisa constate en lisant le document *Statistiques canadiennes sur le cancer 2015*¹³ que le cancer du col utérin représentait 1,5 % des nouveaux cas de cancer et qu'il a été à l'origine d'environ 1,0 % des décès par cancer chez les femmes au Canada en 2015. Pour une femme de l'âge de Jeanne, les chances de survivre pendant 5 ans après un diagnostic de cancer du col utérin sont de l'ordre de 74 %¹⁴. Cependant, la survie à cinq ans varie selon le stade de la maladie au moment de sa découverte. Elle s'élève en effet à environ 90 % chez les personnes présentant une forme localisée de cancer au moment du diagnostic, et elle chute à 40 % dans les cas invasifs¹⁵. Ces observations plaident en faveur de la pertinence du dépistage pour ce type de cancer.

De plus, pour justifier un dépistage, il doit exister un traitement efficace qui, s'il est instauré précocement, pourra améliorer le pronostic. Si la découverte et le traitement précoces de la maladie ne peuvent laisser espérer une réduction de la morbidité et de la mortalité qui y sont associées, la valeur du dépistage peut être mise en doute. Or, Lisa sait que l'incidence des cancers invasifs du col utérin de même que le taux de mortalité lié à cette maladie ont diminué de façon significative depuis l'utilisation de la cytologie cervico-vaginale comme examen de dépistage. Au Canada, l'incidence du cancer du col utérin serait de 7,5 pour 100 000 femmes¹⁶, alors que l'on sait que de 65 à 91 % des femmes selon la catégorie d'âge se soumettent régulièrement aux examens de dépistage¹⁷. En comparaison, par exemple en Afrique subsaharienne où le dépistage précoce du cancer du col utérin n'est pas accessible, le taux d'incidence s'élève à 34,8 cas pour 100 000 femmes¹⁸. Cette observation incite Lisa à croire à l'efficacité de cette activité de dépistage. Des études internationales ont par ailleurs montré l'efficacité des programmes de dépistage bien organisés. Ainsi, on estime que le risque de développement d'un cancer invasif du col utérin chez les femmes ayant subi le dépistage peut être réduit de plus de 90 %. Au Canada, le taux de mortalité attribuable au cancer du col de l'utérus a diminué de 2,3 % par année entre 2001 et 2010. On estime que cette baisse est largement attribuable au dépistage¹⁹. Bien qu'en théorie, le dépistage du cancer du col utérin chez l'ensemble des femmes à risque devrait occasionner une réduction d'au moins 90 % de la mortalité due à cette maladie, la réalité observée est tout autre, principalement parce que divers problèmes sont inhérents au dépistage, tels que la difficulté de rejoindre les populations à risque et le manque de suivi en ce qui a trait aux cytologies jugées anormales. Par exemple, au Canada, environ 74 % des femmes adultes participeraient au dépistage²⁰, alors qu'en France, 40 % des femmes âgées de 25 ans à 65 ans n'auraient jamais passé cet examen²¹. Le cas de Jeanne illustre bien cette lacune, puisque cette dernière a négligé de passer une cytologie de contrôle pendant 20 ans.

Les caractéristiques des examens de dépistage

Il ne suffit pas de tenir compte des caractéristiques de la maladie visée par le dépistage; on doit aussi prendre en considération les particularités de l'examen choisi²². Ainsi, la méthode utilisée doit être reproductible, c'est-à-dire qu'elle doit donner

les mêmes résultats lorsqu'elle est de nouveau employée dans les mêmes conditions sur un même sujet. En d'autres termes, l'examen doit être suffisamment précis pour permettre de distinguer entre un changement significatif du phénomène étudié et des erreurs aléatoires relevant de l'examen lui-même. Dans le dépistage du cancer du col utérin, la technique utilisée pour effectuer la cytologie décrite par Papanicolaou répond à des critères précis, et Lisa sait qu'elle doit les respecter à la lettre pour assurer la précision et la qualité de cet examen. De la même façon, le laboratoire de cytopathologie qui analyse les lames cytologiques doit avoir acquis une expertise dans l'interprétation de celles-ci, ce qui permet une reproductibilité inter-observateurs et intra-observateur adéquate.

L'examen utilisé pour effectuer le dépistage doit également être valide, ce qui signifie que les résultats obtenus doivent permettre de différencier les individus porteurs de la maladie de ceux qui ne le sont pas. La validité de l'examen nécessite donc à la fois une sensibilité et une spécificité élevées. Dans le contexte d'un dépistage, il importe de garder à l'esprit que la prévalence de la maladie à dépister peut être faible. Cela influence à la baisse la valeur prédictive positive* de l'examen de dépistage, même si sa sensibilité et sa spécificité sont élevées. Autrement dit, en présence d'un résultat positif de l'examen de dépistage, il devient essentiel de confirmer la présence de la maladie par des examens complémentaires, car, compte tenu de la faible prévalence de la maladie, la probabilité que le sujet identifié comme positif soit réellement malade peut être relativement faible.

En raison de contraintes méthodologiques, il est difficile d'obtenir des données précises sur la sensibilité et la spécificité de la cytologie du col utérin. Selon la littérature, la spécificité de ce test de dépistage est très élevée, se situant probablement de 90 à 99 %, alors que sa sensibilité se révèle plus basse, peut-être bien autour de 60 à 80 %²³. Cette faible sensibilité nécessite que le test soit répété fréquemment pour garantir une certaine sécurité. Par ailleurs, avec l'avènement de la vaccination contre le VPH, on peut s'attendre à ce que la valeur prédictive positive de la cytologie diminue, compte tenu de la réduction de la prévalence de la maladie²⁴.

Finalement, l'examen utilisé pour le dépistage doit être acceptable pour les sujets qui le subiront; il doit notamment être peu inconfortable et peu effrayant. Ces caractéristiques s'avèrent d'autant plus importantes que le dépistage s'adresse par définition à des femmes asymptomatiques. De la même façon, l'examen doit être simple et facile à exécuter, surtout s'il s'agit d'un dépistage touchant une population nombreuse. Enfin, l'intervention doit engendrer de faibles coûts pour ne pas alourdir de façon inconsidérée les dépenses liées aux soins de santé, ce qui annulerait les avantages que l'on pourrait retirer d'un traitement effectué précocement après un dépistage.

* La valeur prédictive positive est la probabilité que la maladie soit présente lorsque le résultat de l'examen s'avère positif. Cette notion est traitée en détail dans le chapitre 7, qui porte sur les examens diagnostiques.

À retenir

Critères devant être satisfaits avant le dépistage :

- La maladie recherchée doit constituer un problème de santé important.
- L'histoire naturelle de la maladie doit être suffisamment connue.
- La phase de latence ou le stade asymptomatique doit être d'une durée suffisante pour permettre d'effectuer le dépistage.
- Un traitement instauré à ce stade doit pouvoir influencer le pronostic et le cours de la maladie de façon plus favorable que s'il était administré au moment où l'affection devient symptomatique.
- Il doit exister un traitement efficace.
- Il doit exister un examen de dépistage valide, fiable et acceptable aux yeux des patients qui ont à le subir.
- Des critères précis établissant quels sujets feront l'objet d'un diagnostic et d'un traitement doivent avoir été définis à l'avance.
- Le dépistage doit pouvoir se faire de façon continue.

Les conséquences du dépistage

Il importe d'insister sur les conséquences possibles du dépistage, puisque Lisa est l'instigatrice de cette intervention préventive et qu'elle la propose à des patientes comme Jeanne qui sont asymptomatiques sur le plan gynécologique. On s'entend pour reconnaître les retombées favorables du dépistage d'une maladie chez les sujets correctement identifiés comme positifs ou négatifs. Les premiers reçoivent de façon plus rapide et plus précoce un traitement qui a alors plus de chances d'être curatif ou, du moins, de conduire à un meilleur pronostic. De la même façon, les vrais négatifs sont rassurés par le résultat de l'examen. Mais qu'en est-il des gens faussement identifiés comme positifs ou négatifs ?

Les faux positifs sont soumis inutilement à des examens plus poussés, fréquemment plus effrayants et potentiellement iatrogènes, dans le but de confirmer le diagnostic de la maladie. Ces interventions et la menace d'un diagnostic grave alimentent leur inquiétude. Ils peuvent même recevoir des traitements inutiles. En outre, les investigations supplémentaires engendrent des coûts pour la société. Quant aux faux négatifs, rassurés par le résultat du dépistage, ils peuvent se montrer moins attentifs aux signes et aux symptômes précoces de la maladie et tarder indûment à consulter un médecin. Le tableau 16.4 résume les avantages et les inconvénients du dépistage.

En somme, l'enjeu principal du dépistage consiste à reconnaître correctement les sujets réellement malades dans une population asymptomatique, tout en tenant compte de la validité des examens dont on dispose, des ressources disponibles et des contraintes économiques.

Un moyen qui permet de satisfaire à ces exigences est d'établir la périodicité du dépistage. La répétition de l'examen à l'intérieur d'un court intervalle peut augmenter la probabilité de découvrir la maladie, mais elle entraîne des coûts importants. Ainsi,

TABLEAU 16.4 ■ Les avantages et les inconvénients du dépistage

	Maladie +	Maladie –
	Vrais positifs	Faux positifs
Test +	<ul style="list-style-type: none"> • Traitement instauré plus précocement et potentiellement plus efficace • Meilleures chances de survie des patients 	<ul style="list-style-type: none"> • Inquiétude associée à la menace d'un diagnostic grave induite chez les patients • Interventions effractives, potentiellement iatrogènes, pour confirmer ou infirmer le diagnostic
	Faux négatifs	Vrais négatifs
Test –	<ul style="list-style-type: none"> • Faux sentiment de sécurité pouvant rendre les patients moins attentifs à l'apparition de signes et de symptômes, ce qui retarderait l'établissement du diagnostic 	<ul style="list-style-type: none"> • Sentiment de sécurité éprouvé par les patients

il a été estimé que le dépistage du cancer du col utérin, effectué tous les 3 ans chez les femmes de 20 ans à 64 ans, réduirait l'incidence cumulée de cancer invasif de 91 %, tout en nécessitant environ 15 cytologies par femme, ce qui correspond à une fréquence de 96 cas détectés par 100 000 cytologies réalisées. En comparaison, un dépistage du cancer du col utérin répété annuellement réduirait l'incidence cumulée de cancer invasif de 93 %, mais il requerrait près de 45 examens par femme pour une fréquence de seulement 33 cas reconnus par 100 000 examens²⁵.

Les recommandations concernant la périodicité du dépistage peuvent donc varier d'un groupe d'experts à un autre, selon les critères de décision utilisés et les priorités retenues. En ce qui a trait au cancer du col utérin, ces groupes s'entendent pour recommander le dépistage dès l'âge de 21 ans. On préconise le plus souvent, pour des raisons de coûts-avantages, un intervalle de trois ans entre les examens.

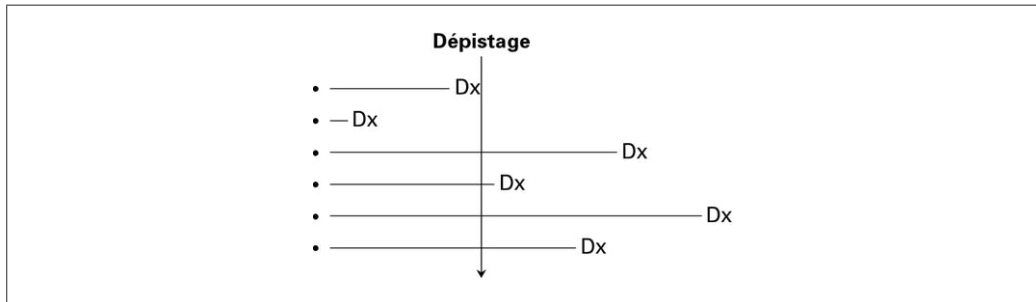
Le biais de durée et le biais de dépassement

À cette étape, il importe justement de considérer certaines des lacunes méthodologiques que comportent les études sur l'évaluation de l'efficacité des activités de dépistage. Ces lacunes, qui peuvent parfois entraîner une fausse impression d'efficacité de l'intervention, sont attribuables au biais de durée — un biais de sélection — et au biais de dépassement — un biais d'information.

Il y a un **biais de durée** lorsque les cas reconnus à la suite d'un dépistage sont d'évolution lente ou sont moins graves que ceux diagnostiqués après l'apparition de symptômes²⁶. En effet, le dépistage permet de désigner les individus atteints d'une maladie alors qu'ils se situent dans une phase asymptomatique. Or, lorsque la maladie évolue lentement, cette phase se trouve en quelque sorte artificiellement prolongée comparativement à ce qui se produit chez les cas à évolution rapide. L'utilisation d'un examen de dépistage tend donc, dans ces circonstances, à privilégier la découverte de cas à évolution lente. Cette situation en surestime l'efficacité et laisse croire à un meilleur

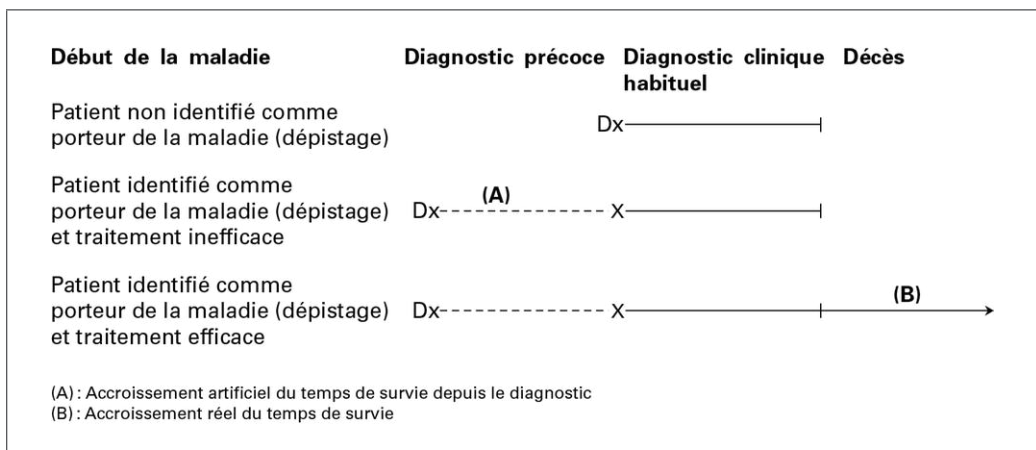
pronostic dans les cas décelés par le dépistage que dans ceux reconnus à la suite d'un diagnostic clinique. Toutefois, cet avantage apparent s'explique plutôt par la sélection préférentielle des cas à évolution lente que par le dépistage lui-même. La figure 16.2 illustre le biais de durée: les cas à évolution lente (espace long entre le début de la maladie [•] et l'apparition des symptômes et l'établissement du diagnostic [Dx]) sont plus susceptibles d'être décelés par le dépistage que ceux qui évoluent rapidement (espace court entre le début de la maladie [•] et l'apparition des symptômes et l'établissement du diagnostic [Dx]).

FIGURE 16.2 ■ Le biais de durée



Un second type d'erreur d'interprétation, appelé **biais de dépassement**, peut survenir au moment de l'évaluation d'une activité de dépistage. Comme cette pratique favorise une découverte plus précoce de la maladie que le permet le diagnostic clinique traditionnel, le temps de survie peut paraître plus long, même si le traitement administré à la suite du dépistage n'améliore pas véritablement le pronostic²⁷. Dans de telles circonstances, il peut sembler que les patients vivent plus longtemps après la découverte de la maladie. En réalité, leur durée de survie n'a pas augmenté; c'est plutôt la période pendant laquelle ils ont eu connaissance de leur état de santé qui a été plus longue. La figure 16.3 illustre le concept de biais de dépassement.

FIGURE 16.3 ■ Le biais de dépassement



Pour contrer les biais de durée et de dépassement, il importe de fonder l'évaluation de l'efficacité des activités de dépistage sur des études de type expérimental. Dans de telles études, la répartition des individus entre les groupes soumis ou non au dépistage se fait par hasardisation, c'est-à-dire de façon aléatoire ; les groupes ont donc toutes les chances d'être comparables quant aux proportions respectives de cas à évolution rapide et de cas à évolution lente. Par ailleurs, l'évaluation de l'efficacité du dépistage par la comparaison des taux de survie entraîne aussi une mesure biaisée (biais de dépassement). En effet, on l'a vu, le dépistage peut mener à une interprétation faussée de la survie. Il est donc recommandé d'évaluer l'efficacité du dépistage par la comparaison des taux spécifiques de mortalité entre le groupe soumis au dépistage et le groupe témoin.

Ainsi, après une analyse des études sur l'efficacité du dépistage du cancer du col utérin, Lisa est rassurée quant à l'absence de tels biais dans celles-ci.

16.3.3 Les considérations éthiques liées au dépistage

Lisa est tenue à la confidentialité en ce qui concerne les informations recueillies auprès de sa patiente ; cette règle s'applique aussi au dépistage. Par ailleurs, d'autres considérations éthiques s'imposent avant d'entreprendre des activités de dépistage, car elles ne sont pas demandées par la patiente au départ. Ainsi, Lisa doit garder à l'esprit le principe de l'autonomie de la personne et de son autodétermination. En d'autres termes, Jeanne a le droit de décider de se soumettre ou non au dépistage recommandé. L'exercice de cette autodétermination exige la capacité de donner un consentement libre et éclairé²⁸. Lisa doit donc fournir à Jeanne toutes les informations nécessaires afin que celle-ci puisse prendre une décision en toute connaissance de cause. Des informations complètes incluent habituellement des renseignements sur la nature de l'examen, sa valeur et ses limites, ses risques et ses conséquences possibles, de même que sur les solutions de rechange offertes. Le langage utilisé doit être accessible et compréhensible par Jeanne. De plus, la patiente doit choisir librement, c'est-à-dire sans crainte de déplaire ou d'être traitée différemment.

Plus précisément, Jeanne doit connaître le risque que la maladie se développe chez elle. Elle doit comprendre que le dépistage effectué pourra mener à d'autres examens, souvent plus effrayants, si le résultat est positif. Enfin, elle doit tenir compte du fait que l'examen de dépistage peut donner des résultats faussement positifs ou faussement négatifs et que ces éventualités comportent des conséquences particulières.

L'exemple du cancer du col utérin illustre bien l'importance d'une analyse rigoureuse des preuves scientifiques avant de conclure à la pertinence d'une intervention préventive. Cette démarche difficile, mais nécessaire, doit être effectuée pour l'ensemble des problèmes de santé pour lesquels il existe des interventions préventives.

16.4 La lecture critique d'une étude portant sur l'évaluation de l'efficacité du dépistage

Au cours de la prochaine rencontre multidisciplinaire à la polyclinique Milo, Lisa doit faire le point sur la pertinence d'offrir la mammographie comme test de dépistage du cancer du sein chez les femmes âgées de 40 ans à 49 ans. Ce sujet a alimenté

beaucoup de controverse depuis plusieurs années. Lisa trouve plusieurs publications à ce propos sur Internet. Elle retient une étude réalisée par Moss et autres²⁹, qu'elle juge de bonne qualité. Plusieurs ECH ont démontré que le dépistage du cancer du sein chez les participantes de plus de 50 ans est associé à une réduction d'environ 25 % de la mortalité par cancer du sein. Par ailleurs, l'efficacité du dépistage chez les femmes qui n'ont pas encore atteint la cinquantaine est moins certaine, bien que les ECH réalisés dans ce groupe d'âge suggèrent une certaine diminution de la mortalité. L'incertitude à cet égard provient du fait que les populations étudiées comprenaient en partie des femmes âgées de plus de 50 ans, ce qui a pu introduire un biais de sélection en faveur de l'efficacité du dépistage.

L'étude de Moss et de ses collaborateurs consiste en un ECH réalisé auprès de 160 921 femmes âgées au départ de 39 ans à 41 ans. Les femmes du groupe expérimental étaient invitées à subir une mammographie annuelle jusqu'à l'âge de 48 ans, et les femmes du groupe témoin à recevoir les soins médicaux habituels. La mesure d'effet est le risque de mortalité par cancer du sein après 10 ans de suivi. Bien que l'étude ait démontré une réduction de la mortalité de 17 % chez les femmes du groupe expérimental, cette diminution n'atteignait pas le seuil de signification statistique.

Pour l'essentiel, la lecture critique d'une étude portant sur l'efficacité du dépistage est soumise aux mêmes exigences que celle d'un ECH; le lecteur peut ainsi se référer à l'encadré 13.2 présenté dans le chapitre 13. Cependant, certains points plus spécifiques relatifs à l'évaluation de l'efficacité du dépistage doivent faire l'objet d'une attention particulière. Parmi ces points supplémentaires figurent notamment la prise en compte des biais potentiels de durée et de dépassement, la description détaillée des avantages et des inconvénients du dépistage, ainsi que la discussion sur la puissance statistique de l'étude, puisque, dans le cas présent, les résultats obtenus n'atteignent pas le seuil préétabli de signification statistique.

16.4.1 La prise en compte des biais potentiels de durée et de dépassement

Comme nous l'avons déjà mentionné auparavant, un ECH est une étude de type expérimental dans laquelle la répartition des individus entre les groupes soumis ou non au dépistage se fait de façon aléatoire, ce qui permet de prévenir le biais de durée, étant donné que ces groupes ont toutes les chances d'être comparables quant à leurs proportions respectives de cas à évolution rapide et de cas à évolution lente.

Nous avons également déjà souligné que l'évaluation de l'efficacité du dépistage par la comparaison des taux de survie entraîne une mesure biaisée (biais de dépassement); il est donc recommandé d'évaluer l'efficacité du dépistage grâce à la comparaison des taux spécifiques de mortalité entre le groupe soumis au dépistage et le groupe témoin. Dans la section «Méthode» de l'étude de Moss et de ses collaborateurs, on indique que l'efficacité du dépistage est évaluée grâce à la comparaison des taux de mortalité totale et de ceux par cancer du sein chez le groupe expérimental et le groupe témoin. Lisa est alors rassurée quant à la prévention et au contrôle de ces biais potentiels dans cette étude.

16.4.2 La description détaillée des avantages et des inconvénients du dépistage

À la section « Discussion », Lisa trouve certains passages où l'auteur discute des inconvénients associés au dépistage du cancer du sein, notamment du risque de cancer induit par les radiations³⁰ :

Dans cette étude, l'utilisation de la mammographie sur un seul plan réduit la dose reçue. La dose moyenne reçue estimée sur un échantillon des centres participants était d'environ 7 % supérieure à la dose reçue chez les femmes de 59 ans et plus, probablement en raison de la plus grande densité du tissu mammaire chez les plus jeunes. Si l'on considère que 5 % des femmes dépistées ont eu une mammographie sur deux plans, le nombre de cancers induits par la radiation par 1 000 participantes âgées de 40 ans à 49 ans est réduit par un facteur d'environ 0,75 [...]. Le pourcentage estimé par cette méthode de femmes pour qui les risques dépassent les avantages serait très petit. [...] Un autre inconvénient concerne la survenue de résultats faussement positifs pouvant provoquer de l'anxiété et nécessitant des investigations plus poussées et possiblement des biopsies inutiles [...]. Au total, 17 030 femmes du groupe expérimental ont accepté toutes les invitations à se présenter pour le dépistage et ont subi au moins 7 mammographies. De ces participantes régulières, 23 % (3 913) ont eu au moins un résultat faussement positif, comparativement à 12 % chez les femmes âgées de plus de 50 ans. De ces 3 913 participantes, 92 % (3 616) n'ont pas eu besoin de subir une cytologie ou une biopsie chirurgicale, 4 % (171) ont eu une cytologie seulement, 2 % (90) ont eu une biopsie et 1 % (36) ont eu une cytologie et une biopsie.

L'auteur traite également des avantages financiers associés au dépistage du cancer du sein chez les femmes plus jeunes³¹ :

L'avantage absolu du dépistage dans ce groupe d'âge en termes de décès prévenus est inférieur à celui des femmes plus âgées, mais le nombre d'années de vie sauvées par décès prévenu a tendance à être plus élevé. Si l'on suppose une moyenne de 35 années sauvées pour chaque décès prévenu et un coût de [...].

16.4.3 La discussion sur la puissance statistique de l'étude

Les résultats de cette étude témoignent d'une réduction de la mortalité chez les femmes du groupe expérimental, cette diminution n'atteignant toutefois pas le seuil de signification statistique. De tels résultats poussent Lisa à s'interroger sur la puissance statistique de l'étude. À la section « Méthode », elle peut lire ce qui suit³² :

L'étude prévoyait à l'origine le recrutement de 190 000 femmes pour détecter une réduction de 20 % de la mortalité après 10 ans de suivi, avec une puissance de 80 % et un seuil de signification statistique de 5 %.

Dans la section « Discussion », Lisa repère le passage suivant³³ :

La puissance de l'étude a été diminuée en raison d'une taille d'échantillon plus petite que prévu et d'un taux de mortalité moins important que celui anticipé dans le groupe témoin, ce qui entraîne une révision de la puissance à 60 % [...].

Après la lecture de l'article, Lisa conclut que même si cette étude présente une bonne validité interne, le manque de puissance statistique ne lui permet pas de conclure quant à la pertinence du dépistage du cancer du sein chez les femmes âgées de 40 ans à 49 ans. Cette étude ne permet pas de renverser le *statu quo*, et la conduite à adopter demeure l'utilisation du jugement clinique et surtout l'obligation de bien informer les patientes des avantages et des inconvénients du dépistage du cancer du sein dans ce groupe d'âge.

En résumé

Les interventions cliniques préventives peuvent se situer à trois niveaux, soit les niveaux primaire, secondaire ou tertiaire, selon la phase de l'évolution de la maladie pendant laquelle elles sont effectuées. Ainsi, la découverte et la réduction des risques pour la santé constituent la prévention primaire. Le dépistage ou la détection précoce des maladies, avant l'apparition de manifestations cliniques, correspond à la prévention secondaire. Enfin, le niveau de prévention tertiaire consiste à intervenir auprès d'individus déjà malades pour contrôler les facteurs modifiables qui ont un impact sur l'évolution de la maladie.

N'étant pas demandées par le patient, les interventions préventives doivent s'appuyer sur des preuves scientifiques et sur la démonstration que l'individu en retire plus d'avantages qu'il ne subit

d'inconvénients. Ces preuves reposent pour l'essentiel sur les résultats d'ECH. Le USPSTF a pour mandat de faire des recommandations sur les services préventifs à intégrer dans la pratique des professionnels de la santé. Ces recommandations sont formulées à partir des preuves scientifiques existantes. Elles tiennent compte du fardeau de souffrance engendré par le problème de santé en cause et de l'efficacité de l'intervention préventive proposée. Des mises à jour de ces recommandations sont publiées régulièrement au fur et à mesure du progrès des connaissances. Les travaux du groupe sont largement fondés sur la reconnaissance de l'importance de la relation entre le médecin et le patient dans l'établissement d'une démarche clinique préventive, et ils s'inscrivent adéquatement dans une perspective de continuité des soins.

Exercices



Voir la section complète des exercices sur le Web.

1. Pour laquelle des raisons suivantes le dépistage du cancer du poumon n'est-il pas recommandé?
 - a) C'est une maladie dont la fréquence ne justifie pas le dépistage.
 - b) C'est une maladie ayant un faible fardeau de souffrance.
 - c) Il n'existe pas d'examen de dépistage valide pour cette maladie.
 - d) L'efficacité des traitements pour ce type de cancer n'est pas connue.
 - e) Le dépistage augmenterait indûment l'anxiété des personnes atteintes.
2. Les moyens recommandés par le USPSTF pour améliorer les interventions préventives sont variés. Lequel de ceux mentionnés ci-après n'en fait pas partie?
 - a) La relation entre le professionnel de la santé et le patient en milieu clinique.
 - b) La recherche de cas (*case finding*).
 - c) Le dépistage de masse.
 - d) L'utilisation d'ensembles de mesures de protection sanitaires.
 - e) La périodicité des interventions préventives.

3. Vrai ou faux. La vaccination des travailleurs de la santé contre l'hépatite B est une mesure de prévention primaire.
4. Vrai ou faux. Le dépistage du cancer du col utérin à l'aide de la cytologie cervico-vaginale est une mesure de prévention secondaire.
5. Vrai ou faux. Le dépistage de masse est un des moyens recommandés par le USPSTF pour améliorer les interventions préventives.
6. Le USPSTF fait des recommandations sur l'efficacité d'une intervention préventive en fonction de la qualité des types d'études recensées dans la littérature médicale. Placez, par ordre de qualité décroissante, ces types d'études?
 - a) L'étude de cohorte(s) ou cas-témoins.
 - b) L'étude expérimentale avec hasardisation.
 - c) Les études descriptives de comparaison entre différents lieux ou différentes périodes temporelles.
 - d) L'étude expérimentale sans hasardisation.
7. Identifiez parmi les biais de sélection suivants celui auquel on doit porter une attention particulière lors de la lecture critique d'un article portant sur le dépistage?
 - a) Le biais d'échantillonnage.
 - b) Le biais de volontariat.
 - c) Le biais de durée.
 - d) Le biais de migration.
8. Identifiez parmi les biais d'information suivants celui auquel on doit porter une attention particulière lors de la lecture critique d'un article portant sur le dépistage?
 - a) Le biais de rappel.
 - b) Le biais de dépassement.
 - c) Le biais de vérification.

Pour les réponses, voir le corrigé à la fin du manuel.

Notes et références

Chapitre 1

1. Audet, N. et H. Leclère. «Les habiletés requises pour la lecture critique en médecine: un cadre de référence issu d'une recension des écrits», *Pédagogie médicale*, 2001, vol. 2, n° 4, p. 206-212, [en ligne], www.pedagogie-medecale.org/articles/pmed/pdf/2001/04/pmed20012p206.pdf (page consultée le 18 mars 2016).
2. Organisation mondiale de la Santé (OMS). *Révision du Règlement sanitaire international*, 2005, [en ligne], www.who.int/csr/ihr/WHA58-fr.pdf (page consultée le 18 mars 2016).
3. Fletcher, R.H., S.W. Fletcher et E.H. Wagner. *Épidémiologie clinique*, Paris, France, Éditions Pradel, 1988.
4. Fonds de la recherche en santé du Québec. *Guide d'éthique de la recherche et d'intégrité scientifique. Standards en éthique de la recherche et en intégrité scientifique du FRSQ*, 2003, [en ligne], www.frqs.gouv.qc.ca/documents/10191/186005/Guide_ethique_integrite_2003.pdf/111344c8-1208-42b5-a0d5-2495f555ed12 (page consultée le 15 juillet 2016).
5. Goldberg, M., Association des épidémiologistes de langue française, Association pour le développement des études et recherches en épidémiologie sur la santé au travail, Association pour l'étude de l'épidémiologie des maladies animales, Association pour le développement de l'épidémiologie de terrain. *Recommandations de déontologie et bonnes pratiques en épidémiologie (version France – 2007)*, [en ligne], www.epiter.org/spip/IMG/pdf/Recommandations_2_-_Version_finale-France-Aout_2007.pdf (page consultée le 15 juillet 2016).
6. Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO). *Déclaration universelle sur la bioéthique et les droits de l'homme*, 2005, [en ligne], unesdoc.unesco.org/images/0014/001461/146180F.pdf (page consultée le 18 mars 2016).
7. Voir la note 6.
8. Voir la note 4.
9. Equator Network. *Translation of reporting guideline*, 2016, [en ligne], www.equator-network.org/library/translations-of-reporting-guidelines (page consultée le 15 juillet 2016).
10. Par exemple :
 - Audet, N. et H. Leclère (voir la note 1) ;
 - Goichot, B. et N. Meyer. *Guide de lecture critique d'un article médical original (LCA)*, Faculté de Médecine de Strasbourg, Strasbourg, France, septembre 2011, [en ligne], http://med.unistra.fr/var/ezwebin_site/storage/original/application/c6b2973ce1a180c73dfcbe597ce2991c.pdf (page consultée le 26 octobre 2016) ;
 - pour les revues systématiques: Moher, D., A. Liberati, J. Tetzlaff et autres. «Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement», *PLoS Medicine*, 2009, vol. 6, n° 7, p. e1000097 ;
 - pour des essais hasardisés: Schulz, K.F., D.G. Altman, D. Moher, for the CONSORT Group. «CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials». *Annals of Internal Medicine*, 2010, vol. 152, n° 11 :726-32 ;
 - pour les études d'observation: von Elm, E., D.G. Altman, M. Egger et autres. «The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies», *Annals of Internal Medicine*, 2007, vol. 147, n° 8, p. 573-577.
11. Voir la note 1.
12. Université de Montréal. *Guide d'utilisation de la base de données en sciences biomédicales MEDLINE, janvier 2016*, [en ligne], <http://guides.bib.umontreal.ca/disciplines/461-MEDLINE> (page consultée le 27 avril 2016); U.S. National Library of Medicine. *MEDLINE/PubMed Resources Guide*, octobre 2015, [en ligne], www.nlm.nih.gov/bsd/pmresources.html (page consultée le 27 avril 2016).

Chapitre 3

1. Ancelle, T. *Statistique Épidémiologie* (3^e éd.), Paris, France, Maloine, 2011.
2. Fletcher, R.H., S.W. Fletcher et E.H. Wagner. *Épidémiologie clinique* (3^e éd.), Paris, France, Éditions Pradel, 1998, p. 243-250.
3. Les lecteurs qui s'intéressent à cette question pourront consulter, pour en savoir davantage, l'ouvrage de Bernard, P.M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Québec, Québec, Presses de l'Université du Québec, 1987.
4. Statistique Canada. *Naissances et décès*, 2016, [en ligne], www.statcan.gc.ca/tables-tableaux/sum-som/l02/cst01/demo07a-fra.htm (page consultée le 26 janvier 2017).
5. Université de Sherbrooke. *Perspective monde. Espérance de vie à la naissance (année) 2014*, [en ligne], <http://perspective.usherbrooke.ca/bilan/stats/0/2014/fr/3/carte/SP.DYN.LE00.IN/x.html> (page consultée le 19 octobre 2016).
6. Institut de la statistique du Québec. *Table de mortalité abrégée selon le sexe, Québec, 2013-2015*, 2015, [en ligne], www.stat.gouv.qc.ca/statistiques/population-demographie/deces-mortalite/307.htm (page consultée le 19 octobre 2016).

Chapitre 4

1. Cet exemple est tiré d'un cas réel rapporté dans Raff, E., P.F. Halloran et C.M. Kjellstrand. «Renal failure after eating “magic” mushrooms», *Journal de l'Association médicale canadienne*, 1992, vol. 147, p. 1339-1341, [en ligne], www.pubmedcentral.nih.gov/picrender.fcgi?artid=1336442&blobtype=pdf (page consultée le 27 octobre 2016).
2. Godlee, F. «Applying research evidence to individual patients. Evidence based case reports will help», *British Medical Journal*, 1998, vol. 316, p. 1621-1622.
3. L'article original est le suivant : Langlois, R. et D. Paquette. «Increased serum lithium levels due to ketorolac therapy», *Journal de l'Association médicale canadienne*, 1994, vol. 150, p. 1455-1456, [en ligne], www.pubmedcentral.nih.gov/picrender.fcgi?artid=1486635&blobtype=pdf (page consultée le 27 octobre 2016). On peut trouver de nombreux autres exemples dans le *Journal of Medical Case Reports* (www.jmedicallcasereports.com), qui se spécialise depuis février 2007 dans la publication de cette forme d'écrits, ainsi que dans la plupart des grands journaux médicaux tels que le *New England Journal of Medicine*, *The Lancet* et le *Journal de l'Association médicale canadienne*.
4. Gagnier, J.J., G. Kienle, D.A. Altman et autres. The CARE Guidelines: Consensus-based Clinical Case Reporting Guideline Development, [en ligne], www.care-statement.org (page consultée le 24 janvier 2017).

Care case report guidelines. *CARE Checklist – 2016: Information to consider when writing a case report*, [en ligne], <http://data.care-statement.org/wp-content/uploads/2016/08/CAREchecklist-English-2016.pdf> (page consultée le 24 janvier 2017).
5. *The CARE Guidelines: Consensus-based Clinical Case Reporting Guideline Development. Flow Diagram – Case Reports and CARE guidelines*, [en ligne], www.equator-network.org/wp-content/uploads/2013/10/CAREFlowDiagram-2013-01F.pdf (page consultée le 13 février 2017).
6. J.H. Kempen. «Appropriate use and reporting of uncontrolled case series in the medical literature», *American Journal of Ophthalmology*, 2011, vol. 151, p. 7-10.e1.
7. Centers for Disease and Control and Prevention. «Pneumocystis pneumonia – Los Angeles», *Morbidity and Mortality Weekly Report*, 1981, vol. 30, p. 250-252, [en ligne], www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm (page consultée le 11 décembre 2016).
8. «Exotic diseases close to home», *The Lancet*, 1999, vol. 354, n° 9186, p. 1221, [en ligne], www.thelancet.com/journals/lancet/article/PIIS0140673699001774/fulltext (page consultée le 27 octobre 2016).
9. Kindhauser, M.K., T. Allen, V. Frank et autres. «Zika: The origin and spread of a mosquito-borne virus», *Bulletin of the World Health Organization*, 2016,

vol. 94, p. 675-686, [en ligne], www.who.int/bulletin/volumes/94/9/16-171082.pdf (page consultée le 27 octobre 2016).

10. Cet exemple est tiré de Creech, J.L. et M.N. Johnson. «Angiosarcoma of liver in the manufacture of polyvinyl chloride», *Journal of Occupational Medicine and Toxicology*, 1974, n° 16, p. 150-151, [en ligne], http://journals.lww.com/joem/Abstract/1974/03000/Angiosarcoma_of_Liver_in_the_Manufacture_of.5.aspx (page consultée le 11 décembre 2016).

Chapitre 5

- Doll, R. et A.B. Hill. «Lung cancer and other causes of death in relation to smoking», *British Medical Journal*, 1956, vol. 2, p. 1071.
- Kahn, H.A. «The Dom Study of smoking and mortality among U.S. veterans: Report on eight and one-half years of observations», dans W. Haenszel (dir.). *Epidemiological approaches to the study of cancer and other chronic diseases*, National Cancer Institute, Monograph 19, U.S. Department of Health, Education and Welfare, Public Health Service, January 1966, p. 1-125.
- Le RRP est le rapport du risque d'un événement dans le groupe considéré, le groupe des exposés (R_g), au risque de cet événement dans la population dont est issu le groupe (R_p). Puisque le groupe fait partie de la population, le R_p est donc la somme du R_g et du risque dans le reste de la population, les non-exposés (R_o). Cette somme sera pondérée par la taille relative du groupe par rapport à la taille de la population (π).

Le R_p , donc, s'exprime comme suit :

$$\pi(R_g) + (1 - \pi)R_o$$

Par conséquent, le RRP s'exprime, quant à lui, de la manière suivante :

$$\frac{R_g}{R_p} = \frac{R_g}{\pi(R_g) + (1 - \pi)R_o}$$

Mais puisque le $RR = R_g/R_o$, alors le RRP est égal à :

$$\frac{RR}{\pi RR + (1 - \pi)}, \text{ soit } \frac{RR}{\pi(RR - 1) + 1}$$

Il ressort clairement de cette dernière relation que si le RR est plus petit que 1, le RRP surestimera le RR. Si le RR est plus grand que 1, le RRP sous-estimera le RR. Dans les deux cas, plus le RR sera près de 1, plus la marge d'erreur sera faible. Par ailleurs, plus petite sera la proportion π , plus bénigne aussi sera l'erreur commise.

4. Formellement, le RRP s'exprime comme suit :

$$R_g/R_p$$

où R_g est le risque de l'événement dans le groupe, et R_p , le risque de l'événement dans la population dont est issu le groupe.

En outre, en supposant que tout le groupe soit à risque au début de l'étude, R_g s'exprime de la manière suivante : $\frac{o(\text{événements observés})}{n_g(\text{taille du groupe})}$

Si A désigne les événements attendus, alors : $A = R_p \times n_g$ et donc : $RRP = o/A$

- Carter, K.C. «Koch's postulates in relation to the work of Jacob Henle and Edwin Klebs», *Medical History*, 1985, vol. 29, p. 353-374.
- Hill, A.B. «The environment and disease: Association or causation?», *Proceedings of the Royal Society of Medicine*, 1965, vol. 58, p. 295-300.
- Ancelle, T. *Statistique Épidémiologie* (3^e éd.), Paris, France, Maloine, 2011; Fletcher, R.H., S.W. Fletcher et E.H. Wagner. *Épidémiologie clinique* (3^e éd.), Paris, France, Éditions Pradel, 1998, p. 243-250; Jenicek, M. *Epidemiology: The logic of modern medicine*, Montréal, Québec, EPI-MED, 1995, p. 166-168; Rothman, K.J. *Modern Epidemiology*, Boston, MA, Little, Brown and Co., 1986, p. 17-20; Rothman, K.J., S. Greenland et L.L. Timothy. *Modern Epidemiology* (3^e éd.), Philadelphie, PA, Lippincott Williams & Wilkins, 2008; Sackett, D.L., R.B. Haynes, G.H. Guyatt et P. Tugwell. *Clinical Epidemiology: A basic science for clinical medicine* (2^e éd.), 1991, p. 294-301.
- Les références suivantes permettent d'approfondir la notion des mesures d'impact :

Ancelle, T. *Statistique Épidémiologie* (3^e éd.), Paris, France, Maloine, 2011; Bernard, P.M. *Analyse des tableaux de contingence en épidémiologie*, Québec, Québec, Presses de l'Université du Québec, 2004, p.159-172; Bernard, P.M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Québec, Québec, Presses de l'Université du Québec, 1987, p. 105-117; Rothman, K.J., S. Greenland et L.L. Timothy. *Modern Epidemiology* (3^e éd.), Philadelphie, PA, Lippincott Williams & Wilkins, 2008.

Chapitre 6

1. Quinn, R.R., B.R. Hemmelgarn, R.S. Padwal et autres. « The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: Part I – blood pressure measurement, diagnosis and assessment of risk », *Canadian Journal of Cardiology*, vol. 26, n° 5, May 2010, p. 241-248.
2. Ancelle, T. *Statistique Épidémiologie* (3^e éd.), Paris, Maloine, 2011, 320 p.; Fletcher, R.H., S.W. Fletcher et E.H. Wagner. *Épidémiologie clinique* (3^e éd.), Paris, France, Éditions Pradel, 1998, p. 23; Scherrer, B. *Biostatistique* (2^e éd.), vol. 1, Montréal, Québec, Gaëtan Morin éditeur, 2007.
3. Bernard, P.-M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Sillery, Québec, Presses de l'Université du Québec, 1987, p. 204.
4. Choi, B.C.K. et A.L. Noseworthy. « Classification, direction and prevention of bias in epidemiologic research », *Journal of Occupational Medicine*, vol. 34, n° 3, 1992, p. 265-271.
5. Checkoway, H., N.E. Pearce et D.J. Crawford-Brown. *Research Method in Occupational Epidemiology*, New York, NY, Oxford University Press, 1989, p. 79; Li, C.Y. et E.C. Sung. « A review of the healthy worker effect in occupational epidemiology », *Journal of Occupational Medicine*, vol. 49, n° 4, 1999, p. 225-229.
6. Turner, E.H., A.M. Matthews, E. Linardatos et autres. « Selective publication of antidepressant trials and its influence on apparent efficacy », *New England Journal of Medicine*, vol. 358, n° 3, 2008, p. 252-260.
7. Piantadosi, S. *Clinical trials: A methodologic perspective*, New York, NY, Wiley-Interscience, 1997, p. 119-120.
8. Il s'agit d'une étude fictive dont le but est de faire comprendre le biais de confusion. Dans la réalité, le risque de séroconversion pour les UDI est beaucoup plus élevé que dans la population en général. Ainsi, on estime à environ 10 à 20 % la prévalence de l'infection chez les UDI au Québec, alors que cette prévalence est de l'ordre de 1 pour 1 000 (0,1 %) dans l'ensemble de la population. On peut consulter à ce sujet: Noël L., J. Laforest et P.R. Allard. *Usage de drogues par injection et interventions visant à réduire la transmission du VIH et du VHC – Revue systématique de la littérature et validation empirique*, Institut national de santé publique du Québec, 2007.
9. Pour une description formelle des méthodes d'ajustement, voir: Bernard, P.-M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Sillery, Québec, Presses de l'Université du Québec, 1987, p. 246-247; Jenicek, M. *Epidemiology: The logic of modern medicine*, Montréal, Québec, EPI-MED, 1995, p. 56-58.

Chapitre 7

1. Perrier, A., P.M. Roy, O. Sanchez et autres. « Multidetector-row computed tomography in suspected pulmonary embolism », *New England Journal of Medicine*, 2005, vol. 352, n° 17, p. 1760-1768.
2. Konstantinides, S.V., A. Torbicki G. Agnelli et autres. « 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism », *European Heart Journal*, 2014, vol. 35, n° 43, p. 3033-3069, 3069a-3069k.
3. Leung, A.A., K. Nerenberg, S.S. Daskalopoulou et autres. « Hypertension Canada's 2016 Canadian Hypertension Education Program Guidelines for blood pressure measurement, diagnosis, assessment of risk, prevention, and treatment of hypertension », *Canadian Journal of Cardiology*, 2016, vol. 32, n° 5, p. 569-588.
4. Torre, L.A., F. Bray, R.L. Siegel et autres. « Global cancer statistics, 2012 », *CA Cancer Journal for Clinicians*, 2015, vol. 65, n° 2, p. 87-108.

5. Vous pouvez consulter le site de la compagnie Qiagen pour en savoir plus sur le test careHPV à l'adresse: www.qiagen.com/ca.
6. Kang, L.N., J. Jeronimo, Y.L. Qiao et autres. «Optimal positive cutoff points for careHPV testing of clinician – and self-collected specimens in primary cervical cancer screening: An analysis from rural China », *Journal of Clinical Microbiology*, 2014, vol. 52, n° 6, p. 1954-.
7. Klok, F.A., I.C. Mos, M. Nijkeuter et autres. «Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism », *Archives of Internal Medicine*, 2008, vol. 168, n° 19, p. 2131-2136.
8. Voir la référence 7.
9. Yaffe, M.-J., N. Mittmann, P. Lee et autres. *Rapports sur la santé. Modélisation de la mammographie de dépistage du cancer du sein dans le contexte canadien : modification et mise à l'essai d'un modèle de microsimulation*, 2015, [en ligne], www.statcan.gc.ca/pub/82-003-x/2015012/article/14293-fra.htm (page consultée le 15 décembre 2016).
10. Keller, T., T. Zeller, D. Peetz, et autres. « Sensitive troponin I assay in early diagnosis of acute myocardial infarction ». *New England Journal of Medicine*, 2009, vol. 361, n° 9, p. 868-877.
11. Sackett, D.L., R.B. Haynes, et P. Tugwell. *Clinical epidemiology. A basic science for clinical medicine* (2^e éd.). Boston, MA : Little, Brown and Company, 1991.
3. Nendaz, M.R. et A. Perrier. «Étude de la validation de la biopsie endobronchique dans le diagnostic de la sarcoïdose. À propos de la place de la biopsie endobronchique dans le diagnostic de la sarcoïdose», *Revue des maladies respiratoires*, 2002, vol. 19, p. 767-777.
4. Bossuyt, P.M., J.B. Reitsma, D.E. Burns et autres. «STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies», *British Medical Journal*, 2015, October 28, vol. 351, p. h5527. doi: 10.1136/bmj.h5527.
5. Kassaï, B. et J. Bacchetta. «Comment lire un article rapportant l'évaluation d'un test diagnostique?» *Médecine*, 2008, vol. 4, n° 2, p. 80-82. doi: 10.1684/med.2008.0237.

En complément d'information: le *Catalogue et index des sites médicaux francophones (CISMef)*, disponible sur Internet à l'adresse <http://doc.cismef.chu-rouen.fr>, permet d'accéder rapidement à plus d'une centaine de lectures critiques d'articles portant sur des examens diagnostiques publiés dans différentes revues scientifiques à l'aide des mots clés de recherche suivants: «lecture critique examen diagnostique» (page consultée le 13 juillet 2016).

Chapitre 8

1. Jaeschke, R., G. Guyatt et D.L. Sackett. «Users' guide to the medical literature», *Journal of the American Medical Association*, 1994, vol. 271, n° 5 p. 389-391 et vol. 271, n° 9, p. 703-707.
2. Schlienger, J.L., B. Goichot et N. Meyer. *Guide de lecture critique d'un article médical original (LCA)*. Strasbourg, France: Faculté de Médecine de Strasbourg, 2011, p. 26, [en ligne], http://med.unistra.fr/var/ezwebin_site/storage/original/application/c6b2973ce1a180c73dfcbe597ce2991c.pdf (page consultée le 8 janvier 2011).

Chapitre 9

1. Agence de santé publique du Canada (ASPC). *Relevé des maladies transmissibles au Canada. Rapport final sur les résultats de la Conférence nationale de concertation sur les maladies évitables par la vaccination au Canada, juin 2005*, mars 2008, vol. 34S2, [en ligne], www.phac-aspc.gc.ca/publicat/ccdr-rmtc/08vol34/34s2/index-fra.php (page consultée le 17 janvier 2017).
2. Ces données sont tirées de: Cazale, L., P. Laprise et V. Nanhou. «Maladies chroniques au Québec et au Canada : évolution récente et comparaisons régionales, Série Enquête sur la santé dans les collectivités canadiennes», *Zoom santé*, Institut de la statistique du Québec, janvier 2009, n° 17.
3. Les définitions retenues pour les effets âge, période et cohorte sont tirées de: Bernard, P.-M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Sillery, Québec, Presses de l'Université du Québec, 1987, p. 11-12.

4. Les extraits présentés sont une traduction et une adaptation libres (par Pierre Deshaies) de : Massé, R., H. Laperrière, H. Rousseau et autres. « *Chlamydia trachomatis* cervical infection : Prevalence and determinants among women presenting for routine gynecologic examination », *Canadian Medical Association Journal*, 1991, vol. 145, n° 8, p. 953-961. Certains passages présentés dans les exemples ne constituent pas des citations intégrales et ont été construits à partir de plusieurs sections de l'article. Ainsi, l'exemple 9.17 est inspiré de la section « Méthode », l'exemple 9.18 est tiré de la section « Résultats », et l'exemple 9.19 est en partie adapté de données tirées de la section « Discussion ». L'Association médicale canadienne se dégage de toute responsabilité en ce qui concerne la traduction et l'adaptation des passages « cités ».
5. Les questions présentées dans la section 9.4 sont largement inspirées de la traduction française de la liste de vérification STROBE publiée par Gedda en 2015, Laboratoire ER3S (Atelier SHERPAS), Unité de recherche pluridisciplinaire Sport, Santé, Société – Université d'Artois, *Kinesither Revue*, 2015, vol. 15, n° 157, p. 34-38, [en ligne], http://ac.els-cdn.com/S1779012314004318/1-s2.0-S1779012314004318-main.pdf?_tid=9faffae8-de5b-11e6-9b16-00000aab0f27&acdnat=1484839758_5cbb96bdb7cfd2900c71090b2e8eb9ca (page consultée le 18 janvier 2017).
6. Agence de santé publique du Canada (ASPC). « Surveillance accrue des méningocoques invasives au Canada, 2006-2011 », *Relevé des maladies transmissibles au Canada*, mai 2014, vol. 40-49.
7. Li, Y., R. Tsang, S. Desai et H. Deehan H. *Relevé des maladies transmissibles au Canada (RMTC). Surveillance accrue des méningocoques invasives au Canada, 2006-2011*, vol. 40-9, 1^{er} mai 2014, [en ligne], <http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/14vol40/dr-rm40-09/dr-rm40-09-surv-fra.php> (page consultée le 13 février 2017).
8. Émond, A. *CADRE D'ORIENTATION pour le développement et l'évolution de la fonction surveillance au Québec*, Québec, Québec, Gouvernement du Québec, Direction générale de la santé publique, mai 2007, p 19-20.
9. Rothman, K.J., S. Greenland et L.L. Timothy. *Modern Epidemiology* (3^e éd.), Philadelphie, Lippincott Williams & Wilkins, 2008, p. 459-460.
10. Institut national de santé publique du Québec. Pageau, M., R. Choinière, M. Ferland et Y. Sauvageau. *Le portrait de santé. Le Québec et ses régions. Édition 2001*, Québec, Les Publications du Québec, 2001, 453 p. ; Institut national de santé publique du Québec en collaboration avec le ministère de la Santé et des Services sociaux du Québec et l'Institut de la statistique du Québec. *Portrait de santé du Québec et de ses régions 2006 : les analyses – Deuxième rapport national sur l'état de santé de la population du Québec*, Québec, 2006, 131 p. ; Ministère de la Santé et des Services sociaux en collaboration avec l'Institut national de santé publique du Québec et l'Institut de la statistique du Québec. *Pour guider l'action – Portrait de santé du Québec et de ses régions*, ministère de la Santé et des Services sociaux, 2011, 156 p., [en ligne], <http://publications.msss.gouv.qc.ca/msss/fichiers/2011/11-228-01F.pdf> (page consultée le 18 juillet 2016).
11. Ministère de la Santé et des Services sociaux en collaboration avec l'Institut national de santé publique du Québec (2016). *La santé de la population : portrait d'une richesse collective*, 86 p., [en ligne], <http://publications.msss.gouv.qc.ca/msss/fichiers/2016/16-228-01.pdf> (page consultée le 2 février 2017).
12. Agence de santé publique du Canada. « Relevé des maladies transmissibles au Canada, Surveillance des maladies – Rôle de l'OMS », décembre 1998, *Relevé épidémiologique hebdomadaire de l'OMS*, vol. 73, n° 43.

Chapitre 10

1. On peut consulter à ce sujet : Bernard, P.-M. et C. Lapointe. *Mesures statistiques en épidémiologie*, Sillery, Québec, Presses de l'Université du Québec, 1987, p. 154-157.
2. La formule permettant de calculer l'aire sous la courbe entre deux points a et b , et donc les proportions, pour une distribution normale est la suivante (où $\pi = 3,1416$, μ = la moyenne de la

variable dans la population et σ = l'écart type de la variable dans la population):

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Chapitre 11

1. L'utilisation du test d'hypothèse exige que le ou les échantillons soient formés aléatoirement (par tirage au sort). Il est à noter que la formation de groupes de comparaison à partir d'un procédé aléatoire (hasardisation), procédure utilisée pour les études de type expérimental, permet également et sans aucune restriction l'application des tests d'hypothèse.
2. Le théorème de la limite centrale permet la transformation Z suivante (voir le chapitre 10, à la page 209, «*La distribution normale centrée et réduite*»):

$$Z = \frac{(89 - 88)}{\left(\frac{6}{\sqrt{40}}\right)}, \text{ soit } 1,05$$

Selon la table de la distribution normale centrée et réduite (voir le tableau 10.4), la probabilité correspondant à $Z = 1,05$ est de 0,1469, soit de 15 %.

3. Le théorème de la limite centrale permet la transformation Z suivante:

$$Z = \frac{(86 - 88)}{\left(\frac{6}{\sqrt{40}}\right)}, \text{ soit } -2,10$$

La probabilité correspondant à $Z = -2,10$ est de 0,0179, soit de 2 %.

4. L'hypothèse alternative est ici une hypothèse bilatérale. Les chercheurs ne présument pas de la nature de la différence; ils désirent vérifier l'hypothèse qui veut que le Novel soit plus efficace que le Périn, mais également celle selon laquelle il serait moins efficace. Les chercheurs croient bien sûr que le Novel est plus efficace que le Périn, mais il s'agit d'un nouveau médicament, et l'on ne peut écarter la possibilité, même faible, que le Novel s'avère moins efficace. Une hypothèse

bilatérale traduit une forme de prudence devant l'inconnu. Une hypothèse est unilatérale lorsqu'elle indique une seule possibilité: par exemple, le Novel est plus efficace que le Périn ($\mu_n < \mu_p$). Cette hypothèse unilatérale ne permet pas de vérifier statistiquement l'hypothèse qui voudrait que le Novel soit moins efficace que le Périn. En règle générale, il est préférable d'utiliser une hypothèse bilatérale lorsqu'on compare deux traitements.

5. La théorie statistique soutient cette intuition. Si H_0 est vraie, on peut construire une distribution d'échantillonnage des écarts entre les valeurs des moyennes de la PAD. Cette distribution est symétrique autour de la moyenne et se rapproche de la distribution normale en fonction de l'augmentation de la taille des échantillons. Cette distribution a une moyenne de 0 et une erreur type calculée à partir des tailles des deux échantillons et de leurs écarts types respectifs. Or, lorsqu'on connaît la moyenne et l'erreur type d'une distribution échantillonnale, il est possible de construire un test statistique. Pour la comparaison de deux moyennes, on utilise le test de Student, aussi appelé test de t , lorsque la taille des échantillons est petite (inférieure à 30 sujets), ou encore le **test du score Z** pour des échantillons dont la taille est égale ou supérieure à 30 sujets.
6. La taille d'un échantillon est calculée à l'aide d'une formule mathématique adaptée à la mesure qui fait l'objet de l'étude (différence de moyennes, différence de proportions, différence de risques, rapport de risques, etc.). Cette formule prend en compte la différence que l'on veut observer, la puissance désirée ($1 - \beta$), mais aussi la variabilité de la mesure et le seuil de signification statistique fixé (α). Ces calculs assurent, dans une certaine mesure, l'obtention d'une puissance acceptable. Cependant, le verdict final repose sur le calcul de la puissance effectué à partir des résultats obtenus dans les échantillons.

Chapitre 12

1. On peut consulter à ce sujet: Rothman, K.J., S. Greenland et T.L. Lash. *Modern Epidemiology* (3^e éd.), Philadelphie, PA, Lippincott Williams & Wilkins, 2008, chapitre 20; Bouyer, J., D. Hémon,

S. Cordier et autres. *Épidémiologie. Principes et méthodes quantitatives*, Paris, France, Les éditions INSERM, 1995, chapitre 12.

2. Voir la note 1.

3. Dagenais, G.R., Z. Ahmed, N.M. Robitaille et autres. «Total and coronary heart disease mortality in relation to major risk factors – Quebec cardiovascular study», *Canadian Journal of Cardiology*, 1990, vol. 6, n° 7, p. 59-65.

Chapitre 13

1. Nordenberg, T. «The healing power of placebos», *FDA Consumer magazine*, 2000, [en ligne], http://permanent.access.gpo.gov/lps1609/www.fda.gov/fdac/features/2000/100_heal.html (page consultée le 21 décembre 2015).

2. Beecher, H.K. «The powerful placebo», *Journal of the American Medical Association*, 1955, vol. 159, p. 1602.

3. Beckett, N.S., R. Peters, A.E. Fletcher et autres. «Treatment of hypertension in patients 80 years of age or older», *New England Journal of Medicine*, 2008, vol. 358, n° 18, p. 1887-1898. doi: 10.1056/NEJMoa0801369.

4. Leveno, K., F.G. Cunningham, S. Nelson et autres. «A prospective comparison of selective and universal electronic fetal monitoring in 34 995 pregnancies», *New England Journal of Medicine*, 1986, vol. 315, n° 10, p. 615-618.

5. Halonen, J., P. Halonen, O. Järvinen et autres. «Les corticostéroïdes pour la prévention de la fibrillation auriculaire après chirurgie cardiaque: un essai contrôlé randomisé», *Journal of the American Medical Association*, 2007, vol. 297, n° 14, p. 1562-1567.

6. Worrall, G.J. «Repairing skin laceration: Does sterile technique matter?», *Le Médecin de famille canadien*, 1987, vol. 3, p. 1185-1187.

7. L'évaluation d'un nouveau traitement se fait habituellement en trois étapes: les études en laboratoire, les études chez l'animal puis les études chez l'humain. Cette dernière étape se déroule en quatre phases. La phase I, qui requiert la participation

de quelques volontaires, sert à ajuster la dose et à étudier la pharmacodynamique d'un médicament. La phase II met à contribution un petit groupe de malades; on s'intéresse alors à l'efficacité thérapeutique et aux effets secondaires du traitement. La phase III constitue l'essai clinique proprement dit. L'étude hasardisée en double aveugle exige la participation d'un plus grand nombre de malades et vise à démontrer la supériorité du nouveau traitement sur le traitement habituel. La phase IV se poursuit après la mise en marché du médicament et cherche à déceler d'éventuels effets indésirables rares et tardifs.

8. The Norwegian Multicenter Study Group. «Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction», *New England Journal of Medicine*, 1981, vol. 304, p. 801-807.

9. James, I.M., D.N. Griffith, R.M. Pearson et P. Newbury. «Effect of oxprenolol on stage-fright in musicians», *Lancet*, 1977, vol. 2, p. 952-954.

10. Moher D., S. Hopewell, K.F. Schulz, V. Montori et autres. Pour le regroupement CONSORT 2010. Explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epi* 2010; vol. 63, n° 8: e1-e37, [en ligne], www.consort-statement.org/Search?q=moher+2001 (page consultée le 17 janvier 2017).

11. Guyatt, G., D. Rennie, M.O. Meade et D.J. Cook. «Therapy», *Users' guides to the medical literature: A manual for evidence-based clinical practice* (5^e éd.). G. Guyatt et D. Rennie (dir), Chicago, AMA Press, 2005, p. 55-79.

12. Labrecque, M., D. Audet, L.G. Latulippe et J. Drouin. Traduction libre de l'article «Homeopathic treatment of plantar warts», *Canadian Medical Association Journal*, 1992, vol. 146, n° 10, p. 1749-1753.

13. Jacobs, J., M. Jiménez, S.S. Gloyd et autres. «Treatment of acute childhood diarrhea with homeopathic medicine: A randomized clinical trial in Nicaragua», *Pediatrics*, 1994, vol. 93, n° 5, p. 719-725.

14. Phillips, B., C. Ball, D. Sackett et autres. *Oxford Centre for Evidence-based – Medicine Levels of Evidence (March 2009)*, [en ligne], www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ (page consultée le 17 janvier 2017).
15. The Norwegian Multicenter Study Group. «Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction», *New England Journal of Medicine*, 1981, vol. 304, p. 801-807.
16. O'Connor, T.Z., T.R. Holford, B.P. Leaderer et autres. «Measurement of exposure to environmental tobacco smoke in pregnant women», *American Journal of Epidemiology*, 1995, vol. 142, n° 12, p. 1315-1321.

Chapitre 14

1. Ghadirian, P., A. Simard et J. Baillargeon. «Tobacco, alcohol, and coffee and cancer of the pancreas. A population-based, case-control study in Québec, Canada», *Cancer*, 1991, vol. 67, n° 10, p. 2664-2670. Les extraits de cet article présentés dans ce chapitre ont été traduits avec la permission de John Wiley & Sons, Inc. Tous droits réservés.
2. Zheng, W., J.K. McLaughlin, G. Gridley et autres. «A cohort study of smoking, alcohol consumption, and dietary factors for pancreatic cancer (United States)», *Cancer Causes and Control*, 1993, vol. 4, p. 477-482.
3. Gedda, M. Traduction française des lignes directrices STROBE pour l'écriture et la lecture des études observationnelles. *Kinesither Rev* 2015;15(157): 34-38. © 2015 Elsevier Masson SAS. Tous droits réservés. Open access under CC BY-NC-ND license, [en ligne], <http://udsmc.u-strasbg.fr/dmg/wp-content/uploads/2015/06/Crite%CC%80res-STROBE-Fr.pdf> (page consultée le 8 février 2017).

Pour la version originale en anglais : von Elm, E., D.G. Altman, M. Egger et autres. *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies, Checklist*, 2007, [en ligne], www.strobe-statement.org/index.php?id=strobe-home (page consultée le 18 juillet 2016).
4. Binstock M., D. Krakow, J. Stamler, J. Reiff et autres. «Coffee and pancreatic cancer : an analysis of international mortality data », *American Journal of Epidemiology*, 1983 November, vol. 118, n° 5, p. 630-640.
5. Spector, T. «Coffee, soya and pancreatic cancer», *The Lancet*, 1981, n° 11, p. 474.
6. Bernarde, M.A. et W. Weiss. «Coffee consumption and pancreatic cancer: temporal and spatial correlation», *British Medical Journal*, 1982, n° 284, p. 400-402.
7. MacMahon, B., S. Yen, D. Trichopoulos et autres. «Coffee and cancer of the pancreas», *New England Journal of Medicine*, 1981, vol. 304, n° 11, p. 630-633.
8. International Agency for Research on Cancer. *IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Volume 51: Coffee, Tea, Mate, Methylxanthines and Methylglyoxal*, Lyon, XOSM, 1991, 513 p. Une mise à jour de ce document se trouve à l'adresse Web suivante : <http://monographs.iarc.fr/ENG/Monographs/vol51/volume51.pdf>.
9. La Vecchia, C. et A. Tavani. «Coffee and cancer risk: an update», *European Journal of Cancer Prevention*, 2007, vol. 16, n° 5, p. 385-389.
10. Guertin K.A., N.D. Freedman, E. Loftfield et autres. «A prospective study of coffee intake and pancreatic cancer : results from the NIH-AARP Diet and Health Study », *British Journal of Cancer*, 2015, vol. 113, n° 7, p. 1081-1085. doi: 10.1038/bjc.2015.235. Epub 2015 September 24.
11. Pour un approfondissement des études cas-témoins et de cohorte et des biais potentiels reliés à chacun de ces devis, on pourra aussi consulter avec intérêt : Hennekens, C.H., J.E. Buring et S.L. Mayrent. *Épidémiologie en médecine*, Paris, Frison-Roche, 1998, 375 p.
12. Voir la note 3.

13. Breslow, N.E. et N.E. Day. *Statistical Methods in Cancer Research, Volume I – The Analysis of Case-Control Studies*, Lyon, IARC Scientific Publications, 1980, n° 32, p. 27.
14. Particulièrement dans les études cas-témoins, il demeure cependant nécessaire de contrôler ces facteurs lors de l'analyse afin d'éliminer le biais de confusion. Pour en savoir davantage à ce sujet, on lira avec intérêt : Rothman, K.J., S. Greenland et L.L. Timothy. *Modern Epidemiology*, 3^e édition, Philadelphie, Lippincott Williams & Wilkins, mars 2008, p. 283-289.
15. Il existe plusieurs façons de définir les niveaux socioéconomiques en utilisant une ou plusieurs des variables suivantes : le revenu, le niveau de scolarité, l'occupation, la zone résidentielle, etc. Puisqu'il existe une corrélation entre plusieurs de ces variables et qu'il est plus facile d'assortir en fonction de la zone résidentielle, Ghadirian et autres ont ici opté pour cette solution en se basant sur l'indicatif téléphonique local. Par la suite, pour plus d'efficacité, ils ont introduit le niveau de scolarité lors de l'analyse par modélisation.
16. Genkinger J.M., R. Li, D. Spiegelman et autres. « Coffee, tea, and sugar-sweetened carbonated soft drink intake and pancreatic cancer risk : a pooled analysis of 14 cohort studies », *Cancer Epidemiol Biomarkers & Prevention*, 2012, vol. 21, n° 2, p. 305-318. doi : 10.1158/1055-9965.EPI-11-0945-T.
17. Freedman N.D., Y. Park, C.C. Abnet et autres. « Association of coffee drinking with total and cause-specific mortality », *New England Journal of Medicine*, 2012, vol. 366, n° 20, p. 1891-1904.
3. Jones, S.M. « Experimentation and innovation in family medicine residency education: The time is now », *Annals of Family Medicine*, 2007, vol. 5, p. 89-90.
4. The Evidence-Based Medicine Working Group. *Users' guides to the medical literature. A manual for evidence-based clinical practice* (2^e éd.). Dans G. Guyatt et D. Rennie (dir.). McGraw-Hill, 2008, 800 p.
5. Mulrow, C.D. « The medical review article: State of the science », *Annals of Internal Medicine*, 1987, vol. 106, p. 485-488.
6. Oxman, A.D. et G.H. Guyatt. « Guidelines for reading literature reviews », *Canadian Medical Association Journal*, 1988, vol. 138, p. 697-703.
7. Greenland, S. et K. O'Rourke. « Meta-Analysis ». Dans K.J. Rothman et S. Greenland (dir.). *Modern Epidemiology*. Philadelphia, PA, Lippincott Williams & Wilkins, 2008, p. 652-682.
8. Sutton, A.J. *Methods for meta-analysis in medical research*. Chichester, Toronto, Ontario, J. Wiley, 2000, xvii, 317 p.
9. Croteau, A. *L'horaire de travail et ses effets sur le résultat de la grossesse: méta-analyse et méta-régression: mise à jour*. Montréal, Québec, Institut national de santé publique du Québec, Direction des risques biologiques environnementaux et occupationnels, 2007, xiii, 149 p.
10. Moher, D., A. Liberati, J. Tetzlaff et autres (2009). « Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement », *PLoS Medicine*, vol. 6, n° 7, e1000097. doi : 10.1371/journal.pmed.100009712.

Chapitre 15

1. Amori, R.E., J. Lau et A.G. Pittas. « Efficacy and safety of incretin therapy in type 2 diabetes: systematic review and meta-analysis », *Journal of the American Medical Association*, 2007, vol. 298, p. 194-206.
2. Egger, M. *Systematic reviews in health care: Meta-analysis in context*. London, Royaume-Uni, British Medical Journal, 2001, xviii, 487 p.
11. DerSimonian, R. et N. Laird. « Meta-analysis in clinical trials », *Control Clinical Trials*, 1986, vol. 7, p. 177-188.
12. Higgins, J.P.T. et S. Green (dir.). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0, mise à jour mars 2011, The Cochrane Collaboration, 2011, [en ligne], www.cochrane-handbook.org. (page consultée le 2 février 2017).

13. Laurin, D. et P.H. Carmichael. «Combining or not combining published results in the presence of heterogeneity? », *American Journal of Clinical Nutrition*, 2010, vol. 92, p.669-70. doi: 10.3945/ajcn.2010.29767.
14. Shea, B.J., J.M. Grimshaw, G.A. Wells et autres. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews, *BMC Medical Research Methodology*, 2007, vol. 7, p. 10. doi:10.1186/1471-2288-7-10.
15. Kung, J., F. Chiappelli, O.O. Cajulis et autres. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic reviews (R-AMSTAR) for grading of clinical relevance, *Open Dentistry Journal*, 2010, vol. 4, p. 84-91. doi: 10.2174/1874210601004020084.
16. Institut national d'excellence en santé et en service sociaux (INESSS). *Documents méthodologiques*, [en ligne], www.inesss.qc.ca/fileadmin/doc/INESSS/DocuMetho/R_Amstar_FR_21012015.pdf (page consulté le 2 février 2017).
17. Higgins, J.P et S.G. Thompson. «Quantifying heterogeneity in a meta-analysis», *Statistics in Medicine*, 2002, vol. 21, p. 1539-1558.
18. The Cochrane Collaboration. *About us*, 2016, [en ligne], www.cochrane.org/about-us (page consultée le 15 février 2017).
19. The Cochrane Library. *Highlighted Reviews*, 2016, [en ligne], www.cochranelibrary.com (page consultée le 15 février 2017).
4. U.S. Department of Health & Human Services, Agency for Healthcare Research and Quality. *U.S. Preventive Services Task Force (USPSTF) : An introduction*, [en ligne], www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/uspstf/index.html (page consultée le 8 février 2016).
5. Voir les notes 3 et 4.
6. U.S. Preventive Services Task Force (USPSTF) *Update on methods: Estimating certainly and magnitude of net benefice*, 2007, [en ligne], www.uspreventiveservicestaskforce.org/Page/Name/update-on-methods-estimating-certainty-and-magnitude-of-net-benefit (page consultée le 20 décembre 2016).
7. Agence de la santé publique du Canada (ASPC). *Cancer du col de l'utérus*, mai 2014, [en ligne], www.phac-aspc.gc.ca/cd-mc/cancer/cancer_du_col_uterus-cervical_cancer-fra.php (page consultée le 18 janvier 2016).
8. Owen, G. «Screening for squamous cell carcinoma of the cervix», *Le médecin de famille canadien*, 1989, vol. 35, p. 1365-1371.
9. Institut national de santé publique du Québec (INSPQ), Direction des risques biologiques et de la santé au travail, Comité sur l'immunisation du Québec (CIQ) et Comité scientifique *ad hoc* VPH. *La vaccination contre les VPH au Québec : mise à jour des connaissances et propositions du comité d'experts*, juillet 2012, [en ligne], www.inspq.qc.ca/pdf/publications/1518_VaccVPHQc_MAJConnPropComiteExperts.pdf (p. iii et 49 consultées le 16 janvier 2016).

Chapitre 16

1. Statistique Canada. *Enquête sur la santé dans les collectivités canadiennes*, [en ligne], www.statcan.gc.ca/tables-tableaux/sum-som/102/cst01/health115b-fra.htm (page consultée le 24 mai 2016).
2. À l'origine, ce groupe était connu sous le nom de Groupe d'étude canadien sur l'examen médical périodique (GECEMP).
3. Canadian Task Force on Preventive Health Care, [en ligne], http://canadiantaskforce.ca/?content=pcp (page consultée le 8 février 2016).
10. U.S. Preventive Services Task Force. *Cervical Cancer: Screening*, mars 2012, [en ligne], www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/cervical-cancer-screening?ds=1&s=cervix cancer (page consultée le 16 janvier 2016).
11. Institut national de santé publique du Québec (INSPQ), Direction des risques biologiques, environnementaux et occupationnels. *Avis sur l'optimisation du dépistage du cancer du col utérin au Québec*, janvier 2009, [en ligne], www.inspq.qc.ca/pdf/publications/915_AvisOptDepCancerCol.pdf (p. 1 consultée le 19 janvier 2016).

12. Organisation mondiale de la Santé (OMS), Centre international de recherche sur le cancer. *Dernières statistiques mondiales sur le cancer. En augmentation à 14,1 millions de nouveaux cas en 2012: L'augmentation marquée du cancer du sein demande des réponses*, Communiqué de presse n° 223, 12 décembre 2013, [en ligne], www.iarc.fr/fr/media-centre/pr/2013/pdfs/pr223_F.pdf (p. 2 consultée le 18 janvier 2016).
13. Comité consultatif de la Société canadienne du cancer. *Statistiques canadiennes sur le cancer 2015*, mai 2015, adaptation en juin 2015, [en ligne], www.cancer.ca/~media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2015-FR.pdf?la=fr-CA (p. 18 et 41 consultées le 18 janvier 2016).
14. Société canadienne du cancer. *Cancer du col de l'utérus. Statistiques de survie pour le cancer du col de l'utérus*, [en ligne], www.cancer.ca/fr-ca/cancer-information/cancer-type/cervical/prognosis-and-survival/survival-statistics/?region=on (page consultée le 18 janvier 2016).
15. Voir la note 14.
16. Voir la note 13, p. 26.
17. Centre de prévention et de contrôle des maladies chroniques, Direction générale de la santé de la population et de la santé publique et Santé Canada. *Rapport d'étape sur la lutte contre le cancer au Canada*, Ottawa, Ontario, Santé Canada, 2004, p. 53-55.
18. Voir la note 12, p. 3.
19. Voir la note 13, p. 44.
20. Voir la note 17.
21. Sancho-Garnier, H. *Traité des infections et pathologies génitales à papillomavirus*, Paris, France, Springer, 2007, p. 145-150.
22. Centers for Disease Control. «A framework for assessing the effectiveness of disease and injury prevention», *Mortality Morbidity Weekly Report*, 1992, p. 41 (RR-3); Wilson, J.M.G. et G. Jungner. *Principles and Practice of Screening for Diseases*, Genève, Suisse, Organisation mondiale de la Santé, 1968.
23. Voir la note 10.
24. Comité sur l'immunisation du Québec (CIQ), Direction Risques biologiques, environnementaux et occupationnels. *Prévention par la vaccination des maladies attribuables aux virus du papillome humain au Québec*, Québec, Québec, Institut national de santé publique du Québec, 2007, p. 6.
25. Dilhuydy, M.H. «Bénéfices et coûts du dépistage de masse du cancer du col utérin», *Gynécologie*, 1987, vol. 38, p. 440-449.
26. Fletcher, R.H., S.W. Fletcher et E.H. Wagner. *Clinical Epidemiology: The Essentials* (2^e éd.), Baltimore, MD, Williams and Wilkins, 1998, p. 161-163.
27. *Ibid.*
28. Lee, J.M. «Screening and informed consent», *New England Journal of Medicine*, 1993, vol. 328, p. 438-439.
29. L'article suivant, dont les passages cités ont été traduits et adaptés par Clément Beaucage, a servi aux fins de l'exemple de cette section: Moss, S.M., H. Cuckle, A. Evans et autres. «Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: A randomised controlled trial», *The Lancet*, 2006, vol. 368, n° 9552, p. 2053-2060.
30. Moss, S.M., H. Cuckle, A. Evans et autres. «Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: A randomised controlled trial», *The Lancet*, 2006, vol. 368, n° 9552, p. 2058.
31. *Ibid.*, p. 2055.
32. *Ibid.*, p. 2055.
33. *Ibid.*, p. 2056.

Corrigé

Chapitre 1

1. a : B ; b : A ; c : C ; d : B ; e : A. ; 2. b ; 3. d

Chapitre 2

1. b et e ; 2. a ; 3. b ; 4. c

Chapitre 3

1. c ; 2. a ; 3. d ; 4. a ; 5. Faux

Chapitre 4

1. Vrai ; 2. c ; 3. Faux ; 4. Faux ; 5. Vrai

Chapitre 5

1. A : c ; B : c ; 2. c ; 3. b ; 4. b

Chapitre 6

1. a ; 2. b ; 3. Une erreur différentielle ; 4. a

Chapitre 7

1. Vrai ; 2. b ; 3. d ; 4. Vrai

Chapitre 8

1. valides, reproductibles, améliorer ; 2. Tous les patients ont été soumis à l'examen de référence ; 3. Faux ; 4. Vrai

Chapitre 9

1. c ; 2. Faux ; 3. d ; 4. fréquence, répartition, personnes, temps, lieu

Chapitre 10

1. a ; 2. b ; 3. b ; 4. Vrai ; 5. 95,4 % ; 6. 50 %

Chapitre 11

1. b ; 2. Vrai ; 3. a ; 4. a ; 5. a

Chapitre 12

1. Vrai ; 2. Vrai ; 3. Vrai ; 4. Vrai ; 5. Faux ; 6. Vrai ; 7. Faux

Chapitre 13

1. e ; 2. b ; 3. b ; 4. c

Chapitre 14

1. a. Faux ; b. Faux ; 2. c ; 3. c ; 4. b

Chapitre 15

1. Faux ; 2. Vrai ; 3. Vrai ; 4. Vrai ; 5. Faux

Chapitre 16

1. c ; 2. c ; 3. Vrai ; 4. Vrai ; 5. Faux ; 6. b, d, a, c ; 7. c ; 8. b

Glossaire

Affectation à l'aveugle (*voir technique en aveugle*)

Ajustement : Ensemble de techniques qui permettent de contrôler un biais de confusion dû à des variables telles que l'âge lorsqu'on compare deux populations. Deux méthodes sont principalement employées : l'ajustement direct et l'ajustement indirect.

Ajustement direct : Méthode d'ajustement dans laquelle la moyenne des taux spécifiques dans la population à l'étude est calculée en utilisant la distribution de l'âge ou d'une autre variable de la population de référence. L'ajustement direct représente ce que le taux brut aurait été dans la population à l'étude si celle-ci avait eu la même distribution que la population de référence pour cette variable.

Ajustement indirect : Méthode d'ajustement dans laquelle la moyenne des taux spécifiques dans la population de référence est calculée en utilisant la distribution de l'âge ou d'une autre variable de la population à l'étude. Le rapport du taux brut dans la population à l'étude comparativement au taux moyen ainsi obtenu représente le risque relatif à la population (RRP).

Analyse de sensibilité : Dans le contexte des articles de synthèse, analyse qui présente des mesures de synthèse pour des sous-ensembles d'études définis par des critères d'inclusion plus sévères. Ce type d'analyse est utile pour explorer les sources d'hétérogénéité liées à certains critères d'inclusion.

Analyse de variances (*voir test de F*)

Analyse multivariée : Ensemble de techniques utilisées lorsque plusieurs variables doivent être étudiées simultanément, en tenant compte de leurs effets les unes sur les autres.

Analyse par sous-groupes : Dans le contexte des articles de synthèse, analyse qui indique si les mesures de synthèse diffèrent selon les strates de variables liées à des caractéristiques de la population ou des études. Ce type d'analyse est utile pour explorer les sources d'hétérogénéité liées à ces variables.

Analyse stratifiée (*voir stratification*)

Appréciation à l'aveugle (*voir technique en aveugle*)

Article de synthèse : Article dans lequel les auteurs font une synthèse et une analyse critique de l'information offerte sur un sujet donné, fondées sur le recensement des études publiées sur ce sujet.

Article de synthèse systématique : Article scientifique dans lequel les auteurs font une synthèse et une analyse critique de l'information offerte sur un sujet donné. Ce type d'article est réalisé selon un protocole préétabli et il se caractérise par l'utilisation d'une stratégie de recherche des études, par des critères explicites lors de la sélection et de l'analyse des études, de même que par la publication de la méthode utilisée à même l'article. Il peut aussi inclure une méta-analyse, soit la synthèse quantitative des résultats.

Assortiment : Méthode de constitution d'un groupe d'étude et d'un groupe de comparaison qui assure une représentation égale des différentes valeurs des variables confondantes dans les deux groupes. L'assortiment vise à réduire au minimum les biais de confusion et à rendre les groupes comparables eu égard aux variables à contrôler.

Biais : Erreur dans les résultats attribuable à une variation systématique ou à une variation aléatoire qui remet en cause la validité de l'étude.

Biais d'admission : Erreur systématique induite dans une étude lorsque les groupes de comparaison (les témoins) sont constitués de patients hospitalisés; de ce fait, les cas et les témoins ne sont pas issus de populations présentant les mêmes risques d'exposition. Aussi appelé «biais de Berkson», le biais d'admission est un biais de sélection.

Biais d'échantillonnage : Erreur systématique induite dans une étude lorsque les personnes asymptomatiques et celles ayant présenté des épisodes de maladies fatales et de courte durée sont moins susceptibles d'être admises dans l'étude. Aussi appelé «biais de survie sélective», le biais d'échantillonnage est un biais de sélection.

Biais d'information : Erreur systématique induite dans une étude lorsque la mesure ou l'observation d'un phénomène est incorrecte.

Biais d'information différentiel (voir erreur de classification différentielle)

Biais d'information non différentiel (voir erreur de classification non différentielle)

Biais de Berkson (voir biais d'admission)

Biais de «bonne santé des travailleurs» : Erreur systématique induite dans une étude lorsque le choix des sujets ou des témoins est effectué parmi une population de travailleurs (personnes détenant un emploi). Le biais de «bonne santé des travailleurs» est un biais de sélection.

Biais de confusion : Erreur systématique induite dans une étude par des variables (variables confondantes ou facteurs de confusion) qui ont un lien tant avec les facteurs d'exposition étudiés qu'avec la maladie observée, sans cependant faire partie de la chaîne causale liant ces deux éléments.

Biais de dépassement : Erreur systématique induite dans une étude, attribuable au fait que la pratique du dépistage permet une reconnaissance plus précoce de la maladie que ne le fait le diagnostic clinique traditionnel. Le temps de survie peut alors paraître plus long, même si le traitement administré à la suite du dépistage n'améliore pas le pronostic. Le biais de dépassement est un biais d'information.

Biais de durée : Erreur systématique induite dans une étude lorsque les cas décelés à la suite d'un dépistage sont moins graves et évoluent moins rapidement que ceux diagnostiqués de façon habituelle, soit après l'apparition de symptômes. Le biais de durée est un biais de sélection.

Biais de migration : Erreur systématique induite dans une étude lorsque des participants à l'étude l'abandonnent avant son terme. Aussi appelé «biais des perdus de vue», le biais de migration est un biais de sélection.

Biais de publication : Erreur systématique susceptible de toucher les articles de synthèse et induite lorsque les études réalisées sur un sujet donné ne sont pas toutes publiées et que, de ce fait, elles ne sont pas accessibles aux fins d'un article de synthèse en particulier. Par exemple, il est possible que des études ne soient pas publiées parce qu'elles présen-

tent des résultats non statistiquement significatifs ; il pourrait s'agir, entre autres, d'études réalisées sur des échantillons de petite taille.

Biais de rappel : Erreur systématique induite dans une étude, attribuable au fait que les personnes présentant un problème de santé peuvent avoir tendance à se rappeler et à signaler les expositions passées à un facteur quelconque de façon différente que le font les personnes en bonne santé. Le biais de rappel est un biais d'information différentiel.

Biais de sélection : Erreur systématique induite dans une étude à cause des méthodes adoptées pour choisir les participants à l'étude.

Biais de vérification : Erreur systématique induite dans une étude lorsque le résultat d'un examen diagnostique influence la décision d'effectuer ou non l'examen de référence. Le biais de vérification est un biais d'information différentiel.

Biais de volontariat : Erreur systématique induite dans une étude lors du recrutement de volontaires pour y participer, les caractéristiques de ces personnes pouvant être différentes de celles des personnes qui décident de ne pas y prendre part. Le biais de volontariat est un biais de sélection.

Cas incident : Nouveau cas d'une maladie observé durant une période donnée.

Cas prévalent : Cas d'une maladie observé à un moment précis.

Causalité : Relation causale (de cause à effet) entre deux phénomènes, appuyée par des raisonnements souvent exprimés sous forme de critères ou de règles. Ainsi, les principaux critères de Hill sont les suivants : la temporalité ; la force de l'association ; la consistance et la validité des résultats ; la spécificité ; la relation dose-réponse (ou dose-effet) ; la vraisemblance ; la cohérence avec les connaissances antérieures ; la plausibilité biologique ; l'analogie.

Coefficient de corrélation (voir coefficient de relation linéaire)

Coefficient de détermination multiple (R^2) : Proportion de la variation de la variable dépendante attribuable à l'ensemble des variables indépendantes comprises dans un modèle.

Coefficient de régression : Mesure indiquant l'amplitude de variation moyenne d'une variable

indépendante lorsque la variable dépendante varie d'une unité.

Coefficient de relation linéaire: Mesure de l'association linéaire entre deux variables. Un coefficient de 1 indique que lorsque la valeur de l'une de ces variables croît, celle de l'autre variable croît également. Un coefficient de - 1 indique que lorsque la valeur de l'une de ces variables décroît, celle de l'autre variable décroît également.

Coefficient de variation: Mesure de dispersion consistant en le rapport de l'écart type à la moyenne, qui s'exprime comme suit :

$$CV = \frac{\sigma}{\mu}$$

Co-intervention: Intervention non prévue au protocole survenant simultanément au traitement étudié au cours d'études expérimentales.

Confusion (voir biais de confusion et variable confondante)

Consensus d'experts: Article de synthèse portant sur un seul sujet et s'appuyant sur des études originales publiées, dont les auteurs sont généralement des groupes d'experts reconnus par leurs pairs. Les conclusions tirées à partir de la synthèse qu'ils effectuent tiennent parfois lieu de guide de pratique pour une spécialité donnée.

Cote: Ratio dont le numérateur et le dénominateur représentent respectivement chacun des deux éléments d'un même ensemble à deux éléments.

Critères d'exclusion: Ensemble de règles ou de conditions qui permettent d'identifier les personnes inaptes à participer à une étude.

Critères d'inclusion: Ensemble de règles ou de conditions qui permettent d'identifier les personnes aptes à participer à une étude.

Dépistage: Détection de maladies à un stade où le patient est encore asymptomatique.

Dépistage de cas: Application de tests (examens) à un individu qui fait partie d'un groupe partageant certaines caractéristiques communes. Le dépistage de cas vise à départager les personnes qui sont probablement porteuses d'une maladie non diagnostiquée, ou d'un facteur de risque d'une maladie, des personnes qui en sont probablement exemptes. Dans le contexte clinique, on parle aussi de «re-

cherche de cas» (*case finding*), qui se traduit par des mesures préventives pour une affection donnée offertes par le clinicien à des patients qui le consultent pour d'autres symptômes.

Dépistage de masse: Mesure de santé publique qui consiste à pratiquer un dépistage à l'échelle d'une population générale.

Diagramme de flux (ou ordinogramme): Représentation graphique du processus de sélection, d'inclusion et d'exclusion ainsi que des motifs d'exclusion et du nombre d'études à chaque étape d'un article de synthèse systématique. Dans les études expérimentales ou semi-expérimentales, il présente, à chaque étape, le recrutement, la hasardisation, le suivi des individus sélectionnés, les nombres d'individus et les critères d'inclusion et d'exclusion.

Distribution asymétrique: Distribution de probabilités unimodale caractérisée par le fait que la médiane, la moyenne et le mode ne coïncident pas. La distribution est dite «asymétrique à droite» (ou «à asymétrie positive») lorsque les valeurs des mesures de tendance centrale sont réparties comme suit: mode < médiane < moyenne.

La distribution est dite «asymétrique à gauche» (ou «à asymétrie négative») lorsque les valeurs des mesures de tendance centrale sont réparties comme suit: moyenne < médiane < mode.

Distribution d'échantillonnage (voir distribution échantillonnale)

Distribution de fréquence: Ensemble des classes d'une échelle de classification avec leur fréquence, souvent présenté sous forme de tableau.

Distribution de probabilités: Modèle de représentation de la fréquence des différentes valeurs d'une variable dans un échantillon ou une population.

Distribution de Z (voir distribution normale centrée et réduite)

Distribution des moyennes échantillonnales: Distribution de probabilités de la moyenne échantillonnale pour tous les échantillons de même taille pouvant être formés à partir d'une population.

Distribution échantillonnale: Distribution de probabilités d'une statistique pour tous les échan-

tillons de même taille pouvant être formés à partir d'une population.

Distribution normale : Distribution de probabilités de la variable lorsqu'elle possède les caractéristiques mathématiques suivantes : être continue, en forme de cloche et symétrique autour de la moyenne (μ). Le premier écart type (σ) correspond à la distance entre la moyenne et le point d'inflexion de la courbe, c'est-à-dire à l'endroit où la courbe, de convexe qu'elle était, devient concave. La moyenne et l'écart type sont les deux paramètres de la distribution normale. Comme la distribution normale est une distribution de probabilités, l'aire sous la courbe est égale à 1.

D'autres caractéristiques de la distribution normale sont également importantes. D'une part, la moyenne, la médiane et le mode ont une même valeur. D'autre part, la moyenne ± 1 écart type contient approximativement 68,3% des événements possibles; la moyenne ± 2 écarts types contient approximativement 95,4% des événements possibles; et la moyenne ± 3 écarts types contient approximativement 99,7% des événements possibles.

Distribution normale centrée et réduite : Distribution normale pour laquelle la moyenne est de 0 et l'écart type est de 1. Elle est aussi appelée « distribution de Z » ou « distribution standard ».

Distribution standard (voir distribution normale centrée et réduite)

Double aveugle (voir technique en aveugle)

Double insu (voir technique en aveugle)

Écart type : Mesure de dispersion exprimant la distance moyenne des valeurs d'une variable par rapport à la moyenne de ces valeurs, qui s'exprime comme suit (dans une population) :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Lorsqu'on calcule l'écart type dans un échantillon, on utilise la formule suivante :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Échantillon : Sous-ensemble de la population à l'étude.

Échantillon électif : Échantillon constitué d'individus sélectionnés par les chercheurs de façon non aléatoire, par exemple en faisant appel à des volontaires.

Échantillonnage aléatoire simple : Méthode d'échantillonnage dans laquelle tous les sujets d'une population ont une chance égale de faire partie de l'échantillon.

Échantillonnage aléatoire stratifié : Méthode d'échantillonnage dans laquelle la population est répartie selon certaines caractéristiques. Toutes les catégories ainsi constituées sont représentées dans l'échantillon. Tous les sujets appartenant à une catégorie donnée ont une chance égale de représenter leur catégorie dans l'échantillon.

Échantillonnage en grappes : Méthode d'échantillonnage dans laquelle la population est répartie en sous-groupes que l'on appelle « grappes ». Les grappes qui font partie de l'échantillon sont généralement désignées au moyen de méthodes d'échantillonnage aléatoire simple ou aléatoire stratifié. Tous les individus qui constituent une grappe ainsi sélectionnée font partie de l'échantillon.

Échantillonnage systématique : Méthode d'échantillonnage aléatoire dans laquelle tous les individus d'une population sont inscrits sur une liste (réelle ou virtuelle). Le premier individu est choisi aléatoirement, et tous les autres le sont ensuite selon une règle fixe.

Échelle nominale : Échelle de classification des valeurs d'une variable dont l'information exprime une qualité ou un attribut et dans laquelle les classes ne peuvent être que nommées (par exemple, le groupe sanguin : A, B, AB, O).

Échelle ordinale : Échelle de classification des valeurs d'une variable dont l'information exprime une qualité ou un attribut, dans laquelle les classes peuvent être nommées et où il existe une relation d'ordre entre les classes (par exemple, le degré de satisfaction : très satisfait, moyennement satisfait, insatisfait).

Échelle par intervalle : Échelle de classification des valeurs d'une variable dont l'information exprime une quantité, dans laquelle les classes peuvent être nommées, et où il existe une relation d'ordre et une relation de distance entre les classes (par exemple, de 1 à 4 ans, de 5 à 9 ans, de 10 à 14 ans, etc.).

Éditorial : Article de synthèse qui reflète le point de vue du comité de rédaction d'un périodique sur un sujet donné.

Effet de l'attention accordée au malade : Effet thérapeutique induit par le simple fait de porter une attention à un malade.

Effet placebo : Effet thérapeutique induit par l'administration d'un placebo.

Endémie : Persistance d'un nombre élevé de cas d'une maladie dans un lieu donné pendant une longue période.

Épidémie : Apparition d'un nombre anormalement élevé de cas d'une maladie selon le lieu et le temps.

Épidémiologie : Science qui s'intéresse aux facteurs qui influencent la fréquence ou la distribution des maladies et des autres phénomènes de santé.

Erreur aléatoire : Erreur attribuable aux fluctuations dues au hasard et qui influe sur la précision des études. Elle entraîne à peu près autant de résultats supérieurs à la valeur réelle recherchée que de résultats inférieurs à cette valeur.

Erreur de classification différentielle : Erreur systématique induite lorsque les participants à une étude sont répartis en deux groupes – exposés et non-exposés (ou malades et non-malades) – au moyen d'instruments de mesure différents selon le groupe.

Erreur de classification non différentielle : Erreur systématique induite lorsque les participants à une étude sont répartis en deux groupes – exposés et non-exposés (ou malades et non-malades) – au moyen d'un même instrument de mesure imparfait ou défectueux.

Erreur de type I : Risque, consenti avant l'étude, de se tromper lorsqu'on rejette l'hypothèse nulle (erreur de type alpha). Le niveau de risque jugé acceptable est fixé par le chercheur et correspond au seuil de signification statistique.

Erreur de type II : Risque de se tromper lorsqu'on retient l'hypothèse nulle alors que c'est l'hypothèse alternative qui est vraie (erreur bêta).

Erreur systématique : Erreur qui remet en cause la validité d'une étude. Ce type d'erreur est aussi appelé « biais ».

Erreur type : Écart type des moyennes échantillonales, qui s'exprime comme suit (où s représente l'écart type dans l'échantillon, et n , la taille de l'échantillon) :

$$\frac{s}{\sqrt{n}}$$

L'erreur type indique avec quelle précision la moyenne échantillonnale estime la moyenne dans la population.

Espérance de vie : Nombre moyen d'années de vie avant le décès des personnes d'un âge donné dans une population, ou âge moyen au décès des personnes de cette population.

Essai clinique (voir étude expérimentale)

Essai thérapeutique (voir étude expérimentale)

Estimation des paramètres : Processus par lequel on suppose que la valeur des paramètres s'apparente à la valeur de la statistique correspondante mesurée dans l'échantillon.

Étendue : Mesure de dispersion consistant en la différence entre la valeur la plus petite et la valeur la plus grande d'un ensemble d'observations.

Étude cas-témoins : Étude à visée étiologique qui implique d'emblée la comparaison de deux groupes : l'un composé de personnes atteintes de la maladie étudiée, appelées les « cas », et l'autre, de personnes n'ayant pas la maladie, appelées les « témoins ». La maladie sert en quelque sorte de point de départ ; on cherche ensuite un ou des facteurs d'exposition antérieurs susceptibles de l'expliquer.

Étude d'un cas : Étude descriptive limitée à l'observation d'un seul cas.

Étude d'une série de cas : Étude descriptive comportant l'observation de plusieurs cas semblables.

Étude de cohorte(s) : Étude longitudinale à visée étiologique ou descriptive d'un ou de plusieurs facteurs susceptibles de provoquer l'apparition d'une maladie dans un groupe (une cohorte) ou des groupes (des cohortes) qui partagent des caractéristiques communes.

Étude de tendance : Étude visant à mesurer la fréquence d'un phénomène (maladie ou décès) et à en évaluer la répartition dans le temps.

Étude descriptive : Étude qui décrit un ou des phénomènes de santé sans chercher à établir leur cause.

Étude descriptive longitudinale : Étude descriptive effectuée sans groupe de comparaison, permettant l'observation des variables d'intérêt pendant une certaine période de temps.

Étude descriptive transversale : Étude descriptive effectuée sans groupe de comparaison et à un moment précis dans le temps.

Étude écologique : Étude dont l'unité d'observation n'est pas l'individu, mais un groupe d'individus tel qu'une classe, une usine, un quartier, une communauté ou une ville.

Étude expérimentale : Étude qui s'apparente aux expériences scientifiques couramment pratiquées en chimie, en physique ou dans les autres sciences de la nature. Toutes les variables de temps, de lieux et de personnes sont contrôlées par les chercheurs, qui peuvent dans ces conditions observer les variations de la maladie en fonction des variations de l'exposition. L'essai clinique et l'essai thérapeutique sont des études expérimentales.

Étude longitudinale : Étude qui se déroule dans le temps et qui permet l'observation de l'évolution des phénomènes étudiés.

Étude non expérimentale : Étude à visée étiologique longitudinale dans laquelle on observe les phénomènes tels qu'ils se produisent naturellement. Ce type d'étude est aussi appelé « étude d'observation ».

Étude prospective : Étude longitudinale, descriptive ou à visée étiologique, dans laquelle le début de la période d'observation est contemporain de la date du début de l'étude.

Étude quasi expérimentale : Étude dans laquelle les chercheurs contrôlent l'ensemble des facteurs, sauf les variables de personnes.

Étude transversale : Étude dans laquelle les données sont recueillies au cours d'une enquête à un moment précis dans le temps, où aucun suivi n'est effectué, et qui permet de décrire un phénomène de santé à un moment donné. Elle produit une photographie de l'état de la situation, au moment de l'étude, dans la population étudiée. La prévalence est la mesure obtenue à la suite de l'étude transversale.

Examen de référence : Mesure étalon qui correspond à la méthode diagnostique la plus valide qu'il soit possible d'utiliser dans une situation clinique donnée. L'examen de référence permet d'obtenir un diagnostic juste et de répartir les individus en groupes de malades ou de non-malades.

Examen diagnostique : Élément de l'anamnèse ou de l'examen physique, ou épreuve biochimique, microbiologique, radiologique ou fonctionnelle qui concourt à un dépistage ou à l'établissement d'un diagnostic.

Examens en parallèle : Utilisation simultanée de plusieurs examens diagnostiques pour obtenir une meilleure sensibilité et une valeur prédictive négative plus élevée.

Examens en série : Suite d'examens diagnostiques au cours de laquelle le résultat du premier examen est un facteur déterminant dans la décision d'effectuer un nouvel examen complémentaire. Cette stratégie d'investigation permet d'obtenir une meilleure spécificité et une valeur prédictive positive plus élevée.

Fraction étiologique chez les exposés : Rapport entre le risque attribuable et le risque chez les sujets exposés. On obtient alors, dans le contexte d'une association positive, la proportion des nouveaux cas de la maladie ou des décès qui, chez les sujets exposés, sont spécifiquement attribuables au facteur d'exposition.

Fraction étiologique totale : Rapport entre le risque chez les sujets exposés et le risque dans la population totale. On obtient alors, dans le contexte d'une association positive, la proportion des nouveaux cas de la maladie ou des décès qui, dans la population totale, sont spécifiquement attribuables au facteur d'exposition.

Fraction prévenue chez les exposés : Rapport entre le risque attribuable et le risque chez les sujets exposés. On obtient alors, dans le contexte d'une association négative, la proportion des nouveaux cas de la maladie ou des décès qui, chez les sujets exposés, ont pu être évités grâce au facteur d'exposition.

Fraction prévenue totale : Rapport entre le risque chez les sujets exposés et le risque dans la population totale. On obtient alors, dans le contexte d'une association négative, la proportion des nouveaux cas de la maladie ou des décès qui, dans la population totale, ont pu être évités grâce au facteur d'exposition.

Guide de pratique : Article rédigé par plusieurs experts qui font une synthèse exhaustive d'un sujet particulier. Les auteurs d'un guide de pratique ont pour mandat d'étudier l'ensemble des preuves scientifiques concernant un sujet donné et de faire, à la lumière de ces preuves, des recommandations à leurs collègues spécialistes.

Hasardisation : Méthode rigoureuse qui consiste à répartir les sujets dans les groupes d'exposés et de non-exposés selon un procédé aléatoire, ce qui signifie que chaque sujet admis à participer à l'étude présente sensiblement la même probabilité de faire partie de l'un ou l'autre groupe. Cette technique permet généralement une répartition plus homogène des variables de personnes dans les groupes.

Hétérogénéité : Variabilité entre les résultats des études qui excède la variabilité attendue, attribuable au hasard.

Histoire naturelle de la maladie : Mode d'évolution spontanée d'une maladie, indépendamment de toute intervention thérapeutique.

Hypothèse alternative : Hypothèse qui constitue l'hypothèse de recherche. Cette hypothèse est désignée par l'abréviation « H_1 ». Dans le test d'hypothèse, sa vraisemblance est confrontée à celle de l'hypothèse nulle, ou « H_0 ».

Hypothèse nulle : Proposition selon laquelle il n'existe pas d'association entre deux variables. Elle est désignée par l'abréviation « H_0 ».

Incidence cumulée : Rapport entre le nombre de nouveaux cas survenus pendant la période d'observation et le nombre de personnes en observation et susceptibles de devenir des cas, au début de l'étude. Il s'agit d'une proportion et d'une mesure du risque, qui doivent toujours être accompagnées de la mention de la durée d'observation.

Indice : Rapport dont le numérateur et le dénominateur n'appartiennent pas au même ensemble.

Inférence : Processus permettant d'induire la valeur d'un paramètre à partir d'une mesure obtenue dans un échantillon représentatif de la population (une statistique).

Intervalle de confiance : Ensemble de valeurs autour de l'estimation d'un paramètre (une statistique) qui ont une probabilité fixée à l'avance de contenir la valeur du paramètre que l'on cherche à estimer.

Justesse d'un examen diagnostique : Proportion des résultats d'un examen diagnostique vrais (vrais positifs + vrais négatifs) par rapport à l'ensemble des résultats obtenus.

Latence (voir phase de latence)

Létalité : Proportion des personnes nouvellement atteintes d'une maladie donnée qui décèdent durant une période de temps donnée. Une personne atteinte d'une maladie peut décéder de cette maladie ou mourir d'une autre cause. La létalité par la cause d'une maladie désigne uniquement la proportion de décès attribuables à cette maladie. La létalité toute cause renvoie à tous les décès survenus durant la période donnée.

Médiane : Mesure dans laquelle on divise en deux parties égales en nombre l'ensemble des valeurs observées, préalablement ordonnées. Elle prend la valeur de l'observation de rang $(n + 1)/2$, où n représente le nombre d'observations. Cela est vrai si n est pair. Si n est impair, deux variables se partagent le centre de la distribution. La médiane correspond alors à la moyenne de ces deux valeurs.

Mesure d'accord Kappa : Mesure du degré d'accord non attribuable au hasard, soit entre différents observateurs (accord interobservateurs), soit chez un même observateur qui apprécie le même phénomène à plusieurs reprises (accord intra-observateur). Le Kappa s'exprime comme suit :

$$\frac{A_o - A}{1 - A_A}$$

où A_o représente l'accord observé, et A_A , l'accord attendu lorsque l'accord est attribuable au hasard.

Mesure d'association : Mesure quantitative qui permet de vérifier l'existence et la force d'une association entre un ou plusieurs facteurs de risque et la survenue d'une maladie ou d'un décès.

Mesure d'association ajustée : Mesure d'association qui tient compte des facteurs de confusion possibles, par le biais, par exemple, de la modélisation.

Mesure d'association brute : Mesure d'association qui ne tient pas compte des facteurs de confusion possibles.

Mesure de dispersion : Mesure qui informe sur la variabilité des observations. L'étendue, les quantiles, la variance, l'écart type et le coefficient de variation en sont des exemples.

Mesure de tendance centrale: Mesure qui se caractérise par le fait que toutes les valeurs observées tendent à se rassembler autour d'elle. La moyenne arithmétique, la médiane et le mode en sont des exemples.

Méta-analyse: Composante d'une synthèse systématique dans laquelle on effectue la synthèse quantitative des résultats de différentes études au moyen de méthodes statistiques appropriées. Elle permet d'augmenter la précision statistique de l'estimation d'un effet et elle facilite l'exploration des sources d'hétérogénéité. L'article de synthèse systématique avec méta-analyse est parfois simplement appelé «méta-analyse».

Méta-régression: Modélisation par régression multivariée où chaque étude est considérée comme un sujet et où chacune des caractéristiques des études est traitée comme une variable du modèle.

Mode: Dans un ensemble de valeurs observées, valeur la plus fréquemment trouvée.

Modèle mathématique: Représentation plus ou moins simplifiée, à l'aide d'équations, de l'association entre des variables.

Modification (ou effet modifiant): Phénomène qui survient lorsqu'un facteur vient augmenter ou diminuer de façon importante l'association entre la variable dépendante et la variable indépendante (lorsqu'on la considère en fonction de différentes strates).

Mortalité maternelle: Indice dont le numérateur représente le nombre des décès de mères avant 10 semaines suivant l'accouchement, et dont le dénominateur correspond au nombre de naissances vivantes. Cet indice est habituellement exprimé pour 100 000 naissances vivantes.

Mortalité néonatale: Proportion des enfants nés vivants et décédés avant 28 jours de vie par rapport à l'ensemble des naissances vivantes. Cette proportion est habituellement exprimée pour 1 000 naissances vivantes.

Moyenne arithmétique: Mesure de tendance centrale obtenue lorsqu'on divise la somme des valeurs observées par le nombre de valeurs observées.

Moyenne géométrique: Mesure de tendance centrale obtenue en extrayant la n^{e} racine du produit des n valeurs observées.

Moyenne pondérée: Mesure de tendance centrale calculée en accordant un poids relatif à chaque élément d'une distribution.

Nomogramme de Fagan: Graphique qui facilite le calcul des valeurs prédictives à partir des rapports de vraisemblance.

Ordinogramme (voir diagramme de flux)

Pandémie: Épidémie qui survient dans une zone géographique très étendue telle qu'un continent ou la planète entière.

Paramètre: Caractéristique mesurée dans une population (moyenne, variance, etc.).

Perdus de vue (voir biais de migration)

Phase d'induction: Période qui s'étend du début de l'exposition à un facteur de risque jusqu'à l'apparition des premières modifications biologiques caractéristiques de la maladie.

Phase de latence (ou latence): Période pendant laquelle la maladie se développe sans que des symptômes se manifestent.

Placebo: Traitement factice ou substance qui n'a aucun effet pharmacologique reconnu.

Plan d'étude en croisé: Dans le cadre d'une étude expérimentale, plan d'étude où tous les sujets qui participent à l'étude reçoivent alternativement le traitement A et le traitement B.

Plan d'étude en parallèle: Dans le cadre d'une étude expérimentale, plan d'étude consistant à comparer deux ou plusieurs groupes simultanément. Après la hasardisation, les patients demeurent dans le même groupe et reçoivent le même traitement tout au long de l'étude.

Population: Ensemble d'éléments (individus ou objets inanimés) dotés de caractéristiques communes.

Population à l'étude: Groupe choisi pour une étude. La population à l'étude provient de la population cible, et elle est celle à partir de laquelle le chercheur sélectionne des éléments pour former un échantillon.

Population cible: Population à laquelle on veut généraliser les résultats d'une étude.

Population dynamique: Population observée pendant une certaine période, qui change avec le temps. On peut y mesurer des taux d'incidence ou des incidences cumulées.

- Population dynamique fermée:** Population dans laquelle tous les sujets sont bien identifiés au début de l'étude, à laquelle personne ne s'ajoute en cours d'étude et où chacun est suivi jusqu'à la survenue de l'événement étudié (maladie, décès, etc.) ou jusqu'à la fin de la période d'observation.
- Population dynamique ouverte:** Population à laquelle des individus peuvent s'ajouter (naissance, immigration, etc.) ou dont ils peuvent se retirer (décès, émigration, déménagement, etc.) durant la période d'observation.
- Population statique:** Population figée dans le temps, à des fins d'observation. Une population statique ne peut donc fournir que des proportions, des ratios ou des indices, et elle ne permet de mesurer ni taux d'incidence ni incidences cumulées.
- Précision:** Caractéristique d'une mesure ou d'une étude correspondant à l'absence relative d'erreur aléatoire. Ainsi, une mesure est considérée comme précise lorsqu'elle fluctue peu.
- Prévalence:** Proportion des personnes affectées par une maladie à un moment donné dans une population.
- Prévention primaire:** Ensemble des mesures visant à diminuer l'apparition de problèmes de santé en réduisant l'exposition aux facteurs de risque lorsque cela est possible.
- Prévention secondaire:** Ensemble des mesures visant à déceler une affection à un stade précoce asymptomatique, le traitement à ce stade devant pouvoir influencer favorablement le pronostic. Ce type de prévention est aussi appelé « dépistage ».
- Prévention tertiaire:** Ensemble des mesures (interventions auprès du patient qui présente des symptômes) visant à freiner l'évolution de la maladie ou à en réduire les conséquences une fois celle-ci établie.
- Probabilité:** Risque ou chance qu'un événement se produise. La probabilité fréquentiste correspond au nombre de fois qu'un événement s'est produit divisé par le nombre total d'événements. La probabilité *a priori* correspond au nombre de fois qu'un événement peut se produire divisé par le nombre total d'événements possibles. Le domaine de variation du calcul d'une probabilité s'étend de 0 (événement impossible) à 1 (événement certain).
- Proportion:** Rapport de deux quantités appartenant au même ensemble et dans lequel le numérateur est inclus dans le dénominateur.
- Puissance de l'étude (voir puissance statistique)**
- Puissance statistique:** Capacité (ou probabilité) d'une étude de déceler, pour une taille d'échantillon donnée, une certaine différence (par exemple, de proportions) entre deux groupes, lorsque celle-ci existe vraiment dans la population. Elle correspond à $1 - \text{erreur bêta}$.
- Quantile:** Mesure de dispersion consistant en la division d'un ensemble de données en un certain nombre de parties égales. Le calcul des quantiles exige que les données soient ordonnées. Lorsqu'on divise le groupe en quatre parties, on obtient des quartiles; lorsqu'on le divise en cinq parties, on obtient des quintiles; etc. Si l'on divise le groupe en 100 parties égales, on obtient des percentiles (ou centiles).
- R² (voir coefficient de détermination multiple)**
- Randomisation (anglicisme; voir hasardisation)**
- Rapport:** Expression la plus générale de la relation entre deux quantités qui peuvent indifféremment appartenir ou non au même ensemble.
- Rapport de cotes (RC):** Rapport de la cote de l'exposition chez les cas sur la cote de l'exposition chez les témoins.
- Rapport de vraisemblance:** Indice qui correspond, pour un résultat donné d'un examen diagnostique, au rapport de la probabilité d'être malade sur la probabilité de ne pas l'être.
- Ratio:** Rapport dans lequel le numérateur et le dénominateur appartiennent au même ensemble, mais où le numérateur n'appartient pas au dénominateur.
- Régression à effet proportionnel:** Technique d'analyse multivariée estimant le risque que survienne une maladie ou un événement, en tenant compte du temps écoulé entre l'exposition à la variable indépendante et la survenue de la maladie ou de l'événement, et permettant de calculer des rapports de taux d'incidence.
- Régression de Cox (voir régression à effet proportionnel)**
- Régression linéaire:** Méthode statistique permettant d'établir la relation entre une variable dépendante et

une variable indépendante, utilisée lorsque la variable dépendante est une variable quantitative continue. Elle mesure l'amplitude du changement de la variable dépendante pour chaque unité de la variable indépendante.

Régression linéaire multiple : Technique d'analyse multivariée qui permet d'analyser l'association entre une variable dépendante et plusieurs variables indépendantes lorsque la variable dépendante est une variable quantitative continue.

Régression logistique : Technique d'analyse multivariée utilisée lorsque la mesure d'association est un rapport de cotes et lorsque la variable dépendante est catégorielle et dichotomique (c'est-à-dire ne comportant que deux valeurs possibles).

Relation dose-effet (ou dose-réponse) : Association entre un facteur d'exposition et une maladie pour laquelle le risque de la contracter varie en fonction de l'intensité ou de la durée de l'exposition au facteur.

Restriction : Stratégie de contrôle du biais de confusion utilisée au moment de l'échantillonnage, au moyen de laquelle la variable confondante est restreinte à certaines catégories.

Risque attribuable (RA) : Différence entre le risque de contracter une maladie donnée pour les individus exposés et le risque correspondant pour les individus non exposés. Le risque attribuable est une mesure d'impact.

Risque relatif (RR) : Risque de survenue d'une maladie donnée parmi les individus exposés à un facteur donné, divisé par le risque correspondant parmi les individus non exposés. Le risque relatif est une mesure d'association.

Risque relatif ajusté : Résumé des risques relatifs spécifiques obtenus pour chacune des strates au cours d'une analyse stratifiée.

Risque relatif à la population (RRP) : Rapport du nombre de cas observés dans la population à l'étude (un groupe) sur le nombre de cas attendus si la population à l'étude avait le même taux de mortalité ou de morbidité que la population de référence (la population de référence étant le plus souvent la population générale, de laquelle est issue la population à l'étude). En anglais, cette mesure d'association porte le nom de *Standardized Mortality (or Morbidity) Ratio* (SMR). La notion de risque relatif à la

population est une extension de la notion de risque relatif.

Risque relatif brut : Mesure d'association entre un facteur d'exposition et une maladie, obtenue en faisant le rapport du risque d'être malade chez les sujets exposés au risque d'être malade chez les sujets non exposés. Le risque relatif brut est une mesure non ajustée.

Score de propension : Variable créée dans un modèle d'analyse multivariée pour remplacer par une seule variable un ensemble de variables indépendantes.

Score Z : Opération qui permet de transformer les distributions normales en une distribution normale centrée et réduite, et qui s'effectue comme suit :

$$\frac{x - \mu}{\sigma}$$

Sensibilité : Probabilité que le résultat d'un examen soit positif chez un individu malade.

Seuil de positivité : Seuil à partir duquel le résultat d'un examen diagnostique est considéré comme anormal. Ce seuil est aussi appelé «seuil de normalité».

Seuil de signification statistique (voir erreur de type I)

Signification clinique : Caractéristique d'un résultat qui porte à conséquence sur le plan clinique. Il s'agit d'un jugement clinique, et non pas statistique.

Signification statistique : Caractéristique d'un résultat dont la probabilité de l'observer par hasard (valeur p) est égale au seuil de signification fixé ou plus petite que celui-ci.

Simple aveugle (voir technique en aveugle)

Simple insu (voir technique en aveugle)

SMR (voir risque relatif à la population)

Spécificité : Probabilité que le résultat d'un examen diagnostique soit négatif chez un individu non malade.

Statistique : Valeur mesurée dans un échantillon. Cette valeur peut être la moyenne, l'écart type, la médiane, etc.

Stratification : Méthode de contrôle d'un facteur de confusion donné lors de l'analyse, qui consiste à répartir les données dans des strates en fonction de

la présence ou de l'absence du facteur de confusion. Elle permet de présenter des mesures d'association spécifiques.

Synthèse qualitative: Article scientifique qui regroupe, présente et discute les résultats de plusieurs études qualitatives sur un même objet.

Synthèse quantitative (voir méta-analyse)

Synthèse systématique (voir article de synthèse systématique)

Taux: Rapport qui mesure la vitesse de survenue d'un événement dans une population.

Taux d'attaque (TA): Proportion des personnes malades par rapport aux personnes exposées à un facteur de risque reconnu pendant la période d'observation. Il s'agit d'une mesure d'incidence cumulée utilisée surtout en infectiologie.

Taux d'incidence: Mesure de la vitesse de propagation d'une maladie ou d'un phénomène de santé. On trouve au numérateur les événements comptabilisés représentant le nombre de nouveaux cas survenus depuis le début de l'étude jusqu'à sa fin. Au dénominateur, on trouve le cumul du temps écoulé entre le début et la fin de l'observation de chaque personne participant à l'étude et susceptible (à risque) de contracter la maladie étudiée: il s'agit en fait du cumul du temps observé avant la survenue de chaque nouveau cas ou avant la fin de la période d'observation. L'observation prend fin quand la personne tombe malade ou quand l'étude se termine.

Taux de mortalité: Taux d'incidence des décès dans une population.

Technique en aveugle: Technique consistant à déterminer si les sujets d'une étude feront partie du groupe recevant le traitement étudié ou du groupe ne le recevant pas en faisant:

- une affectation à l'aveugle: ni le chercheur ni le patient ne savent à quel groupe le patient a été affecté;
- une appréciation à l'aveugle: la personne qui apprécie ou mesure les résultats ignore quel traitement le patient a reçu.

Lorsque l'une ou l'autre de ces techniques est utilisée, on dit que l'on mène une étude en «simple aveugle» ou «à simple insu». L'étude est dite en «double aveugle», ou «à double insu», lorsqu'elle

comprend à la fois une affectation et une appréciation à l'aveugle. L'étude est dite «en triple aveugle», ou «à triple insu», lorsqu'il y a en plus un comité indépendant de contrôle des données qui effectue l'évaluation des résultats sans connaître l'affectation des sujets.

Test d'hypothèse: Méthode d'inférence statistique qui permet de choisir, non sans risque d'erreur, entre deux hypothèses statistiques, soit l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1). Après avoir formulé les hypothèses statistiques, on utilise un procédé qui vise à convaincre, au moyen de la réfutation de l'hypothèse nulle, de la vraisemblance de l'hypothèse alternative. Pour y parvenir, on suppose que l'hypothèse nulle est vraie. Si cette supposition ne résiste pas à l'examen des données observées, elle est rejetée au profit de l'hypothèse alternative.

Test de t : Analyse statistique utilisée en remplacement du test de Z lorsque la taille des échantillons est inférieure à 30.

Test du khi carré (χ^2): Analyse statistique qui permet de déterminer si la différence entre deux ou plusieurs proportions indépendantes est significative sur le plan statistique.

Test du score Z : Analyse statistique qui permet de déterminer si l'écart entre deux moyennes ou entre deux valeurs d'une variable est significatif sur le plan statistique.

Théorème de la limite centrale: Lorsque, dans une population, on choisit de façon aléatoire tous les échantillons possibles de même taille, on observe que:

- la moyenne de la distribution échantillonnale, c'est-à-dire la moyenne des moyennes échantillonnales, est égale à la moyenne de la variable dans la population si la distribution de cette variable est normale;
- l'erreur type est l'écart type des moyennes échantillonnales et peut être estimée par (s/\sqrt{n}) ;
- pour une taille d'échantillon suffisamment grande ($n \geq 30$), la distribution des moyennes échantillonnales est normale, quelle que soit la distribution de la variable dans la population.

Valeur: Tout état que peut prendre une variable étudiée.

Valeur p : Dans le cadre d'un test d'hypothèse, valeur qui mesure la vraisemblance de l'hypothèse nulle et

qui correspond à la probabilité d'obtenir par hasard un résultat aussi extrême ou plus extrême que celui qui a été observé.

Valeur prédictive négative (VPN): Probabilité qu'une personne ayant obtenu un résultat négatif lors d'un examen diagnostique ne soit pas malade.

Valeur prédictive positive (VPP): Probabilité qu'une personne ayant obtenu un résultat positif lors d'un examen diagnostique soit réellement malade.

Validité: Caractéristique d'une mesure ou d'une étude correspondant à sa capacité d'appréhender correctement un phénomène ou sa valeur réelle.

Validité externe: Capacité de généraliser un résultat (obtenu dans le cadre d'une étude) à une population cible.

Validité interne: Capacité d'une étude d'estimer correctement les mesures ou les associations recherchées.

Validité intrinsèque: Capacité d'un examen diagnostique de reconnaître les malades et les non-malades, caractérisée par deux paramètres: la sensibilité et la spécificité de cet examen diagnostique.

Variable: Tout facteur susceptible d'être différent selon les personnes, le temps ou les lieux considérés.

Variable confondante (voir biais de confusion)

Variable de lieux: Attributs géographiques tels que le pays, la région, la zone urbaine ou rurale, etc. Les variables de lieux permettent de répondre à la question: « Où sont les malades? »

Variable de personnes: Attributs anatomiques, physiologiques, sociaux ou culturels. Les plus fréquemment considérés sont l'âge, le sexe, l'état civil, les habitudes de vie, l'occupation, le statut socioéconomique et le niveau de scolarité. Les variables de personnes permettent de répondre à la question: « Qui est malade? »

Variable dépendante: Variable dont la valeur dépend de l'effet ou de l'influence des autres variables (soit les variables indépendantes). Elle représente le phénomène ou l'événement étudié.

Variable de temps: Attributs temporels tels que l'année, la saison, etc. De façon générale, la fréquence

de la maladie varie avec le temps. La variable de temps permet de répondre à la question: « Quand les personnes sont-elles devenues malades? »

Variable indépendante: Variable qui affecte ou influence le phénomène ou l'événement étudié (soit la variable dépendante).

Variable qualitative (catégorielle): Variable qui exprime une caractéristique de personne, de lieux ou de temps selon une qualité, un attribut.

Variable quantitative: Variable qui exprime une caractéristique de personne, de lieux ou de temps selon une quantité, à l'aide d'un nombre.

Variable quantitative continue: Variable qui exprime une caractéristique de personne, de lieux ou de temps selon une quantité, à l'aide d'un nombre et qui peut être représentée par une fraction.

Variable quantitative discrète: Variable qui exprime une caractéristique de personne, de lieux ou de temps selon une quantité, à l'aide d'un nombre et qui ne peut pas être représentée par une fraction.

Variance: Mesure de dispersion consistant en la somme des carrés des écarts par rapport à la moyenne divisée par le nombre d'observations.

Variations attribuables à l'examen: Sources d'erreurs systématiques ou aléatoires (ou les deux) qui peuvent influencer le résultat d'un examen diagnostique telles que l'environnement dans lequel est réalisé l'examen, l'interaction entre l'examineur et la personne examinée, l'utilisation inadéquate des outils diagnostiques et l'emploi d'appareils défectueux ou mal calibrés.

Variations attribuables à l'examineur: Sources d'erreurs systématiques ou aléatoires (ou les deux) qui peuvent influencer le résultat d'un examen diagnostique telles que les variations biologiques, l'interprétation de la réalité, la classification des diagnostics, les attentes de l'examineur et les lacunes dans les connaissances de ce dernier.

Variations attribuables à la personne examinée: Sources d'erreurs systématiques ou aléatoires (ou les deux) qui peuvent influencer le résultat d'un examen diagnostique telles que les variations biologiques, les effets de la maladie ou du médicament administré et les souvenirs.

Index

A

Accord
inter-observateur, 163
intra-observateur, 163
Affectation à l'aveugle, 65
Ajustement, 60, 124
direct, 61
indirect, 61
Analogie, 101
Analyse(s)
de sensibilité, 275, 276, 330
multivariée, 6, 245-247,
250-257, 306
par sous-groupes, 280, 329
plan d', 281, 282, 305-307
stratifiée, 175, 249, 279-281
univariée, 243-245
utilité, 245-250
Anamnèse, 5, 6, 76, 117
Anomalies, distribution des, 156
Appréciation à l'aveugle, 65
Article de synthèse, 322, 324
classification et fréquence
des, 323
systématique, 324-326
critique à l'aide de
critères objectifs,
332-341
modification de sa pra-
tique suite à la lecture
d'un, 341, 342
Association, 99
force de l', 101
mesures d', 88, 93-95, 233,
245-250, 254, 328,
330, 331
Assortiment, 123, 306, 307
Avantages
de l'examen diagnostique, 162
Aveugle
affectation à l', 65
double, 119
simple, 119
technique en, 156, 157,
277, 278
triple, 119

B

Biais (*voir aussi* Erreurs systé-
matiques), 108, 298-
304, 311, 312, 330
d'admission, 113
d'attention portée aux cas
et aux témoins, 304

d'échantillonnage, 112,
113
d'information, 117, 119,
120, 155, 302, 303
prévention des, 277, 278
réduction des effets du,
119
de Berkson, 113
de « bonne santé des tra-
vailleurs », 115
de confusion, 120, 124, 188,
246-248, 250, 251
contrôle du, 123, 248
prévention des, 278-281
de dépassement, 357-360
de durée, 357-360
de mémoire, 303
de mesure de l'exposition
et de la maladie, 302
de migration, 113, 114
de publication, 115, 327,
330, 331
de rappel, 118, 149, 304
de refus de participer, 298,
311
prévention, 299
de sélection, 111, 112,
116, 175, 296-302,
311, 357
contrôle des, 275-277
réduction des effets du,
116
de source d'information
pour mesurer la
maladie, 303
de surveillance médicale,
302
de survie sélective, 112
de vérification, 157, 158
de volontariat, 114, 116
des décès avant ou durant
l'étude, 301
des perdus de vue, 113, 116
différentiel, 303, 304
non différentiel, 303, 304

C

Cas
-témoins, étude, 63, 66, 67,
92, 95, 112, 113, 247,
288-291, 294-296,
300-305, 317, 318
description d'un, 76
étude d'un, 62, 68, 74,
75, 80, 81

Causalité, 98-102
Chronologie, 81
Classes
collectivement exhaustives,
21
mutuellement exclusives,
21
Classification
des études épidémiolo-
giques, 64-70
des recommandations, 20-23
des recommandations
émises par le
USPSTF, 349
des résultats d'un examen
diagnostique, 128, 129
échelle de, 20, 21
erreurs de, 118
Co-interventions, 270
Coefficient
de corrélation linéaire, 95-98
de près de 0, 97, 98
négatif, 97
positif, 97
de détermination multiple
(R^2), 252
de régression, 251, 252
de variation, 39
Cohorte,
étude de, 62, 66, 122, 255,
289, 291-293, 309,
310
Collaboration Cochrane, 342
Combinaisons d'examens,
143-147
en parallèle, 143-145
en série, 145-147
Compétence
à critiquer une publica-
tion scientifique en
sciences de la santé,
13, 14
Conclusion d'une publication
scientifique, 12, 271,
282-284, 308, 314
Confiance, intervalle de, 207,
221, 232-234, 240
Confusion
biais de, 120, 124, 188,
246-248, 250, 251
Consensus d'experts, 322
Consentement libre et éclairé,
11, 82, 359
Conséquence, 99
Considérations éthiques, 10

Contrôle
STROBE, 291-294, 308
R-AMSTAR, 332, 333, 343
Cote(s), 51
rapport de, 92, 93, 205
Courbe récepteur opérateur
(ROC), 136
Critère(s)
d'exclusion, 40, 41, 275,
296, 331
d'inclusion, 40, 41, 263,
264, 274, 327, 331
de comparaison de l'examen
diagnostique, 155
de Hill, 100
de la causalité, 99-101
objectifs, 332-341
pour juger de la validité d'une
étude d'un examen
diagnostique, 153-155

D

Déclaration financière d'une
publication scienti-
fique, 284, 294
Dépistage, 347, 353, 354
avantages et inconvénients
du, 357, 361
conséquences du, 356, 357
considérations éthiques
du, 359
évaluation de l'efficacité
du, 359-362
Description d'un cas, 76
Diagramme
de flux, 283, 327
en barres, 25
proportionnelles, 24
en secteurs proportionnels,
24
en tartes, 24
Différences
inter-observateurs, 147
intra-observateurs, 147
Dimension chronologique, 76
Discussion dans une publication
scientifique, 12, 77, 81,
82, 271, 282-284, 293,
308, 314
Dispersion, mesures de, 36-39,
81
Distance-temps, 47
Distribution
asymétrique, 32

- d'échantillonnage, 212, 222, 225, 239, 240
- de fréquences, 23, 24
- représentations graphiques, 24-30
- de probabilités, 207
- de Z, 209
- des anomalies, 156
- des moyennes échantillonnelles, 212
- des sujets anormaux, 156 normaux, 156
- échantillonnage, 212, 214 normale, 208-213, 215, 238-241 centrée et réduite, 209, 238 standard, 209
- Données descriptives, 307, 313 obtenues, 307, 313
- E**
- EBM (*Evidence-based medicine*), voir Médecine basée sur les faits
- Écart type, 38, 39, 81, 215
- Échantillon(s), 39, 40, 203-206, 214-216, 271 aléatoire, 263 de convenance, 264 électif, 205, 264
- Échantillonnage, 39-42, 221-224 aléatoire simple, 41 stratifié, 41 biais d', 112, 113, 116 distribution d', 212, 222, 225, 239, 240 en grappes, 42 systématique, 42
- Échelle(s) de classification, 20, 21 nominale, 21, 22, 24, 128 ordinale, 21, 22, 24, 128 par intervalle, 22, 25, 128
- Éditorial, 322
- Effectif, 23
- Effet ampleur de l', 270 de l'attention accordée au malade, 260, 261 placebo, 260 réel d'un traitement ou d'une intervention, 261 taille de l', 283
- Énoncé
- CONSORT, 269
- PRISMA, 327
- Épidémie, 196
- Épidémiologie apport de l', 3-9 définition, 4 perspective clinique, 5, 6 perspective interdisciplinaire, 6-10 perspective populationnelle, 4, 5 principes éthiques, 10
- Erreur(s) aléatoires, 109, 111 de classification, 118 différentielle, 118 non-différentielle, 118 de type I, 229 de type II, 228, 229 systématiques (*voir aussi* Biais), 108, 110, 112, 117, 149, 247 type, 214, 222
- Espérance de vie, 57, 58, 59
- Essai(s) clinique, 6, 10 considérations éthiques particulières aux, 266 hasardisés, 353, 360, 362 communautaire, 262 thérapeutique, 6, 10 hasardisés, 64
- Étendue, 36, 81
- Étude(s) à visée étiologique, 64-66, 84, 92, 95, 120, 288-291 cas-témoin, 63, 66, 92, 92, 95, 247 de type expérimental, 64, 65 cas-témoins, 63, 66, 67, 92, 95, 112, 113, 247, 288-291, 294-296, 300-305, 317, 318 analyse critique, 291-309 considérations éthiques, 317, 318 en milieu hospitalier, 289, 301 imbriquée, 289 populationnelle, 289, 301 d'un cas, 74, 75 avantages et limites, 75, 83 enjeux éthiques, 84 d'une série de cas, 68, 78-81 avantages et limites, 80, 83 caractéristiques, 78 enjeux éthiques, 84 de cohorte, 62, 66, 122, 255, 289, 291-293, 309, 310 analyse critique, 291, 309-316 considérations éthiques, 317, 318 expérimentale, 64, 65 historique, 62, 289 non expérimentale, 70 prospective, 62, 63, 289 quasi expérimentale, 70 de tendance, 69, 179-182 descriptive, 67-69, 119, 124 de population longitudinale, 67 de population transversale, 67 écologique, 68, 288 en double aveugle, 65 en simple aveugle, 65 épidémiologiques, 7, 8, 63 classification des, 64-70 expérimentale, 64, 120, 124, 220, 261-263, 266 analyse critique, 269-284 hasardisée, 246, 249, 262 non hasardisée, 262 menée en croisé, 268 menée en parallèle, 267 longitudinales, 62, 64 non expérimentale, 65 objectifs de l', 273, 274 puissance de l', 230 quasi expérimentale, 262, 263 taille de l', 304, 305, 313 transversales, 61, 64 vérification de la validité interne de l', 155-158
- Évaluation diagnostique, 82
- Examen(s) clinique, 76, 82 de dépistage, 354, 355 de référence, 128, 129, 155, 157, 158 diagnostique, 128, 129, 153-155, 159, 160, 162, 163 application des résultats de l', 162, 163 comparaison avec un examen reconnu, 155 critères permettant l'évaluation de la validité d'un, 153-155 sources de variation dans les, 147-150 utilité de l'_ pour améliorer la qualité des soins, 162, 163 en parallèle, 143-145 en série, 145-147 sensible, 132 spécifique, 132
- Exclusion, critères d', 40, 41, 275, 296, 331
- F**
- Facteurs de risque, 3, 5, 74, 80, 100, 120
- Fagan, nomogramme de, 169, 170
- Fardeau de souffrance, 347, 349, 353
- Fraction étiologique chez les exposés, 102 totale, 103 prévenue chez les exposés, 103 totale, 104
- Fréquence(s) absolue, 23 distribution de, 23 mesures de, 45-49, 174, 175 polygone de, 26
- G**
- GECSSP (Groupe d'étude canadien sur les soins de santé préventifs), 346, 351
- Graphe de percentiles, 27, 28 en lignes, 26, 27 en points, 29, 30
- Grille d'analyse d'un publication portant sur une étude expérimentale, 270-272

- d'analyse des études
 - cas-témoins et de cohorte(s), 291-294
- d'analyse des études d'un cas, 81, 82
- d'analyse des études d'une série de cas, 81, 82
- Guides de pratique, 322
- H**
- Hasardisation, 124, 220, 265, 279
- Hétérogénéité, 322, 329
- Histogramme, 25, 26
- Histoire naturelle de la maladie, 174
- Hypothèse, 262
 - alternative, 224-226
 - bilatérale, 226
 - nulle, 224
 - test de, 207, 221, 224-230, 237-239
- I**
- Incidence(s)
 - cumulée(s), 50, 52, 53, 90, 234
 - taux d', 52, 90, 91, 255-257
- Inclusion, critères d', 40, 41, 263, 264, 274, 327, 331
- Indicateur(s), 55
- Indice(s), 48
- Inférence, 206, 219
 - statistique, 190, 220, 221
- Information(s)
 - additionnelle, 82
 - sur le patient, 81, 82
- Interprétation et « généralisabilité » des résultats, 316, 330, 331
- Intervalle(s) de confiance, 207, 221, 232-234, 240
 - comparaison d', 232, 233
 - et les mesures d'association, 233, 234
 - vérification d'une hypothèse à l'aide d'une, 232-234
- Intervention, 82
 - effet réel d'une, 261
 - facteurs expliquant la réponse à l', 259-261
 - précoce et aide aux personnes atteintes, 5
 - préventive, 347, 350
 - catégories d', 350
- niveaux de certitude du bénéfice réel, 347, 348
- Issue des soins, 76, 77, 82
- K**
- Kappa, 147, 148, 278, 281
- L**
- Lecture critique, 11-15
 - cadre de référence, 13-15
 - habiletés requises pour la, 11, 12
- Létalité, 56, 175, 178, 179
 - taux de, 56
- M**
- Maladie(s)
 - dépistables, 353, 354
 - facteurs de risque de la, 60
 - histoire naturelle de la, 174, 259, 261, 315, 351-353, 356
- Médecine basée sur les faits, 269, 342
- Médiane, 33, 34, 81
- Mesure(s)
 - brute, 122
 - d'accord Kappa, 147, 148
 - d'association, 88, 93-95, 233, 245-250, 254, 328, 330, 331
 - ajustée, 245, 246, 254
 - brute, 245, 246
 - modification de la, 248-250
 - d'intérêt, 251
 - de dispersion, 36-39
 - de fréquence, 45-49
 - de l'impact d'un facteur sur la fréquence des maladies et des décès, 102-104
 - de mortalité et morbidité, 54-59
 - de tendance centrale, 30, 36
 - en épidémiologie, 49-54
 - synthèse des, 52
- Méta-analyse, 322-330
- Méta-régression, 327, 329, 330
- Méthode
 - d'une publication scientifique, 12, 81, 271, 274, 291, 295, 310, 311
 - de reproduction d'un examen, 159
- Mode, 35, 81
- Modèle
 - à effets aléatoires, 328, 329
 - à effets fixes, 328
 - mathématique, 124, 250
 - ou (type d'étude) en
 - épidémiologie, 61-63
- Modification, 248-250
- Morbidité, 56, 57
- Mortalité, 54-56
 - taux de, 54, 55, 90, 91, 178, 353, 354, 360
- Motif de consultation, 76
- Mots clés, 81
- Moyenne(s), 81, 213, 214
 - arithmétique, 31
 - échantillonnales, 212
 - géométrique, 32, 33
 - pondérée, 31, 32
- N**
- Négatif(s)
 - faux, 129, 356
 - vrai, 129, 356
- Niveaux de certitude concernant le bénéfice réel de l'intervention préventive, 347, 348
- Nombre de personnes à traiter (NPT), 104, 270
- Nomogramme de Fagan, 169, 170
- Normalité
 - clinique, 216
 - statistique, 216
- O**
- Ordinogramme, 283
- P**
- P, valeur 226, 227
- Pandémie, 196
- Paramètre(s), 202, 203, 206, 221, 223, 231, 232
- Percentiles
 - graphe de, 27, 28
- Perdus de vue, 274-277
 - biais des, 113, 114, 116
- Personnes à traiter, nombre de, 104
- Personnes-temps, 47
- Phase
 - d'induction, 352
 - de latence, 352
- Placebo, 260
 - effet, 260
- Plan
 - d'analyse, 281, 282
 - d'étude, 266-269
 - en croisé, 267, 269
 - en parallèle, 266, 269
- Plausibilité biologique, 101
- Point de vue du patient, 82
- Polygone de fréquences, 26
- Population, 202, 271, 296-298, 307
 - à l'étude, 112, 113, 204, 301, 309-311
 - cible, 108, 113, 204, 220, 231, 263, 264, 317
 - dynamique, 53
 - fermée, 53, 54
 - ouverte, 53, 54, 55
 - statique, 52
- Positif(s), 129
 - faux, 129, 356
 - vrai, 129, 356
- Postulats de Koch, 99, 100
- Pratiques cliniques préventives, 346-351
- Précision, 40, 109-110, 232, 281
- Prévalence, 49, 90, 138-141
- Prévention, 275-277, 351-359
 - clinique, 345, 346
 - rôle du personnel en, 345, 346
 - de la maladie et des décès, 5
 - primaire, 352
 - secondaire, 352
 - tertiaire, 352
- Principes éthiques de la recherche et de l'intervention en santé, 10, 11
- Prise en charge diagnostique et thérapeutique, 76
- Probabilité(s), 199
 - a priori*, 200
 - fréquentiste, 200
 - propriété d'une, 201, 202
- Proportion(s), 46, 50-52, 131, 135, 136, 138
 - rapport de, 90
- Publication scientifique, 269
 - de synthèse, 322, 323
 - évaluation de la qualité portant sur l'étude d'un cas, 75
 - évaluation de la qualité portant sur l'étude d'une série de cas, 80, 81
 - structure d'une, 12

Puissance
de l'étude, 230
statistique, 40, 230, 231,
281, 361, 362

Q

Quantiles, 37

R

R^2 (coefficient de détermination multiple), 252
Raisonnement diagnostique et thérapeutique, 76
Randomisation, *voir* Hasardisation
Rapport(s), 46
de cotes (RC), 92, 93, 205
de proportions, 90
de taux, 90, 91
de vraisemblance, 161, 166
Ratio(s), 48
RC, *voir* Rapport(s) de cote
Recommandation clinique, 317
Refus de participer, 274-276,
298-300, 311
Régression
à effet proportionnel, 250,
255-257
coefficient de, 251, 252
de Cox, 250, 255-257
linéaire, 244, 245
multiple, 251-253
simple, 251
logistique, 250, 253-255
Relation dose-réponse, 100, 101
Représentation cartographique, 29
Reproductibilité, 147, 159,
160
Restriction, 123, 280, 306
Résultats, 160, 161, 313
d'une publication scientifique, 12, 81, 271,
274, 293, 294, 307
du test d'hypothèse, 227,
228
généralisation des, 270
interprétation et «généralisabilité» des, 316
Résumé d'une publication
scientifique, 12, 75, 80,
81, 270, 291, 309, 310
Risque
attribuable (RA), 88, 89
facteurs de, 3, 5, 74, 79,
351, 352
relatif (RR), 89-91, 205

à la population (RRP),
94, 95
ajusté, 246
brut, 121, 246

S

Score
de propension, 250
Sélection
aléatoire, 263, 289
biais de, 111, 112, 116,
175, 296-302, 311,
357
des sujets, 112, 116, 263,
264
Sensibilité, 130-132, 134-140,
142-145, 157, 158,
160, 161, 169, 188,
275, 355
Série de cas
étude d'une, 78-81
Seuil
de positivité, 132-135, 168
de signification statistique,
227, 229, 238, 239,
330, 360, 361
Signification
clinique, 219, 234-236
statistique, 220-224, 235
seuil de, 227, 229, 238,
239, 330, 360, 361
SMR, *voir* Risque relatif à la
population
Sources de variations du
résultat d'un examen
diagnostique, 108,
109, 147-150
Spécificité, 101, 130, 131,
142, 157
Statistique, 205, 206
Stratification, 124
Suivi des soins, 76, 77, 82
Sujet(s)
effet de l'attention accordée
au, 261, 262
répartition par hasardisation,
265
sélection des, 112, 116,
263, 264
volontaire, 114, 264
Synthèse
qualitative, 115, 328
quantitative, 115, 322,
324-328, 330
systématique, 322-327,
330-332, 334, 336,
341-342
étapes de réalisation,
326-331

T

Taille de l'étude, 304, 305
Taux, 46
d'attaque (TA), 56
d'incidence, 50-54, 90, 91,
255-257
de létalité, 56
de mortalité, 54, 55, 90, 91,
178, 353, 354, 360
Technique en aveugle, 156,
157, 277, 278
Temporalité, 99, 101
Tendance
centrale
mesure de, 30, 36
études(s) de, 69
Test
d'hypothèse, 207, 221,
224-230
de t , 239, 339
du khi carré, 240, 339
Théorème de la limite centrale,
212, 215, 221, 240
Titre d'une publication
scientifique, 75, 80,
81, 291, 294, 309, 310
Traitement
effet réel d'un, 261
facteurs expliquant la
réponse à, 259-261
U
USPSTF (U.S. Preventive
Services Task Force),
346-351, 353, 362
classification des recom-
mandations émises
par le, 349
stratégie d'utilisation des
interventions préven-
tives, 350
V

V

Valeur(s), 17
de référence, 17
extrême, 33
normale, 17
nulle, 90
 p , 226, 227
prédictive, 138, 139, 142,
169
négative (VPN), 137, 138
positive (VPP), 137,
138
Validité, 107, 108, 110, 111,
129

d'une étude portant
sur un examen
diagnostique,
153-155, 160, 161
des résultats, 269
du questionnaire, 312
externe, 108, 129
évaluation de la, 281,
283, 284
extrinsèque, 137, 143
interne, 108, 155-158
évaluation de la, 274,
275, 282, 283
intrinsèque, 130-136, 140,
160, 161
prédictive, 160, 161
Variable(s), 17-19, 214, 298,
311
catégorielle, 19, 95
dichotomique, 253
classification des, 18, 19
confondante(s), 247, 248,
256, 257, 279, 290
continue, 19, 21, 26, 208
de lieux, 18, 66, 177
de personnes, 18, 66
de temps, 18, 64, 174,
176
dépendante, 243-248,
250-255
discrète, 19, 31
indépendante, 244, 245,
247, 248, 251-253,
255, 257
qualitative, 19
quantitative(s), 19, 244,
253, 305, 313
continue(s), 95
discrète, 19
Variance, 37-39, 329
Variation(s)
attribuables
à l'examen, 147, 149,
150
à l'examineur, 147,
148, 150
à la personne examinée,
147, 149, 150
coefficient de, 39
sources de, 108, 109,
147-150
Vérification, biais de, 157, 158
Vraisemblance, 101, 161

Z

distribution de, 209
score, 210, 226, 238-241

En ce XXI^e siècle, une meilleure gestion des ressources dans le secteur de la santé et des services sociaux s'impose toujours comme une priorité à la fois politique, économique et sociale. Ce mouvement de société conduit le domaine de la santé à repenser son rôle et à revoir constamment ses façons de faire. De cette inévitable réflexion est née une obligation : distinguer les connaissances essentielles et nécessaires de celles qui sont secondaires ou désuètes, et appuyer les décisions et les actions des intervenants du milieu sur des faits.

L'objectif premier de cet ouvrage consiste à fournir à l'étudiant et à tout professionnel de la santé les outils nécessaires à une pratique éclairée. Porter un regard critique sur la littérature scientifique s'avère une clé indispensable à l'atteinte de cet objectif. Pour ce faire, il est capital de posséder des connaissances minimales en biostatistique et en épidémiologie, sciences fondamentales de la recherche dans le domaine de la santé.

Un ouvrage de référence complet et indispensable

Seul ouvrage sur le sujet écrit en français par une équipe multidisciplinaire composée d'auteurs représentant l'ensemble de la communauté des sciences de la santé, ce manuel ne se limite pas aux notions théoriques de base. Il ouvre la porte à une application clinique grâce à une approche axée sur le raisonnement critique plutôt que sur la seule mémorisation.

- Les connaissances présentées sont adaptées aussi bien à une clientèle universitaire qu'à des cliniciens et à d'autres professionnels de la santé.
- De multiples exemples concrets d'application des notions d'épidémiologie et de biostatistique sont donnés dans les différentes disciplines du domaine de la santé.
- Des grilles d'analyse critique des divers types d'études sont fournies.
- Chaque chapitre comporte un tableau qui présente les habiletés à acquérir, les objectifs d'apprentissage et leur contenu associé.

Les uns y trouveront un outil d'autoapprentissage sur mesure. Les autres y verront un précieux ouvrage de référence qui répondra à leurs interrogations survenant en contexte clinique.

Ce volume, devenu un classique, intègre les commentaires et les suggestions issus de 20 années d'enseignement universitaire. Sa troisième édition est enrichie afin de répondre aux nouveaux besoins de la formation en la matière.

Un classique dorénavant numérique

L'ouvrage *Épidémiologie appliquée* est associé à un site Web, lequel offre, entre autres, des exercices accompagnés de la mention de leur objectif, les corrigés de ces exercices ainsi que des hyperliens d'intérêt. Le volume, également disponible en version numérique, est aussi adapté aux technologies actuelles et à la formation en ligne.

À la 1^{re} édition, l'ouvrage a remporté :

- le Prix de la Ministre 1998 Volume et Qualité du français Universitaire
- le Prix Prescrire du livre médical et pharmaceutique 1997

Et à la 2^e édition, il a obtenu :

- le Prix d'excellence en enseignement 2009 Volume Université Laval



Les directeurs de l'ouvrage

André Simpson
Clément Beaucage
Yv Bonnier Viger

Découvrez les ressources* en ligne Interactif

ÉTUDIANT

- Le manuel en version numérique
- Des exercices supplémentaires pour chacun des chapitres
- Une annexe sur les méthodes qualitatives
- Une liste d'hyperliens

ENSEIGNANT

- Les figures, tableaux et encadrés de l'ouvrage
- Le corrigé des exercices supplémentaires

*Matériel sujet à changement sans préavis

