

LIGNES DIRECTRICES POUR LA PRÉVENTION ET LE CONTRÔLE DES INFECTIONS :

TROUSSE D'OUTILS DE L'ÉVALUATION CRITIQUE

PROTÉGER LES CANADIENS CONTRE LES MALADIES



PROMOUVOIR ET PROTÉGER LA SANTÉ DES CANADIENS GRÂCE AU LEADERSHIP, AUX PARTENARIATS, À L'INNOVATION ET AUX INTERVENTIONS EN MATIÈRE DE SANTÉ PUBLIQUE.

— Agence de la santé publique du Canada

Also available in English under the title:

Infection Prevention and Control Guidelines Critical Appraisal Tool Kit

Pour obtenir des copies supplémentaires, veuillez communiquer avec :

ccdic-clmti@phac-aspc.gc.ca

On peut obtenir, sur demande, la présente publication en formats de substitution.

© Sa Majesté la Reine du chef du Canada, représentée par la ministre de la Santé, 2014

Date de publication : Juillet 2014

La présente publication peut être reproduite sans autorisation pour usage personnel ou interne seulement, dans la mesure où la source est indiquée en entier. Toutefois, la reproduction en multiples exemplaires de cette publication, en tout ou en partie, à des fins commerciales ou de redistribution est interdite sans l'autorisation écrite préalable du ministre de Travaux publics et Services gouvernementaux Canada, Ottawa (Ontario) K1A 0S5 ou copyright.droitdauteur@pwgsc.gc.ca.

Cat. : HP40-119/2014F-PDF

ISBN : 978-0-660-22556-2

Pub. : 140183

L'Agence de la santé publique du Canada (Agence) élabore des lignes directrices sur la prévention et le contrôle des infections, qui présentent des recommandations fondées sur des preuves en complément aux efforts des autorités provinciales et territoriales en matière de santé publique quant à la surveillance, à la prévention et au contrôle des infections liées aux soins de santé.

L'objet du présent document, *Trousse d'outils de l'évaluation critique*, est de fournir un outil d'évaluation des données probantes, qui éclaire les recommandations présentées dans le cadre de la série de Lignes directrices de prévention et contrôle des infections.

La *Trousse d'outils de l'évaluation critique* a été élaborée par une équipe composée des employés de l'Agence et un critique de Cochrane avec une expertise en méthodologie. Cette équipe se rapportait au Groupe de travail d'experts en prévention et en contrôle des infections (anciennement le Comité directeur chargé de l'élaboration des lignes directrices pour la prévention et le contrôle des infections). Voir l'annexe C pour obtenir la liste des membres du comité.

Les renseignements figurant dans le présent document étaient à jour au moment de la publication. Des recherches et des révisions peuvent s'imposer pour tenir compte de l'évolution ou des changements dans la façon d'aborder une évaluation critique.

TABLE DES MATIÈRES

| | |
|---|----|
| Partie 1 : Consignes et définitions | 2 |
| Partie 2 : Outils pour nommer le plan d'étude, classer les données probantes et rédiger des recommandations | 8 |
| Partie 3 : Outils d'évaluation critique | 29 |
| Dictionnaire de l'outil d'évaluation critique d'une étude analytique | 30 |
| Outil d'évaluation critique d'une étude analytique | 45 |
| Dictionnaire de l'outil d'évaluation critique d'une étude descriptive | 50 |
| Outil d'évaluation critique d'une étude descriptive | 60 |
| Dictionnaire de l'outil d'évaluation critique d'une revue de littérature | 63 |
| Outil d'évaluation critique d'une revue de littérature | 76 |
| Annexe A : Glossaire, liste des abréviations et synthèse des tests statistiques communs | 80 |
| Annexe B : Exemple d'un tableau de synthèse des données probantes et recommandations | 92 |
| Annexe C : Membres du Groupe de travail d'experts en prévention et en contrôle des infections | 95 |
| Bibliographie | 97 |

Liste des tableaux

| | |
|--|----|
| Tableau 1 – Définition des termes utilisés pour évaluer les données probantes | 6 |
| Tableau 2 – Plans d'étude analytique et descriptive | 21 |
| Tableau 3 – Contenu pertinent pour un tableau de synthèse des données probantes | 25 |
| Tableau 4 – Critères pour classer les données probantes à la source des recommandations | 28 |
| Tableau 5 – Synthèse des tests statistiques communs | 90 |
| Tableau 6 – Exemple d'un tableau de synthèse des données probantes et de recommandations | 92 |

Liste des figures

| | |
|---|----|
| Figure 1 : Algorithme pour sélectionner le bon outil | 11 |
| Figure 2 : Algorithme pour nommer le type d'étude analytique | 14 |
| Figure 3 : Algorithme pour nommer le type d'étude descriptive | 17 |
| Figure 4 : Algorithme pour nommer le type de revue de littérature | 20 |

PARTIE 1 : CONSIGNES ET DÉFINITIONS

INTRODUCTION

La présente trousse d'outils a été créée pour l'évaluation critique de la littérature scientifique. Les groupes de travail chargés de l'élaboration de lignes directrices peuvent utiliser l'outil pour promouvoir l'uniformité dans l'évaluation d'un corpus de données probantes, le classement des données probantes et l'élaboration de recommandations connexes. Bien que la présente trousse d'outils soit utile pour évaluer des études préliminaires aux fins de la recherche, rédiger des articles d'analyse et élaborer des politiques, elle ne fournit pas de directives sur la façon de réaliser une revue de littérature. La trousse d'outils comprend ce qui suit :

Système de classement des données probantes et définitions

1. Cinq ensembles d'outils :
2. Outils pour nommer le plan d'étude (algorithmes assortis de légendes)
 - a) Consignes pour la rédaction de tableaux de synthèse des données probantes et de recommandations
 - b) Étude analytique : Dictionnaire de l'outil d'évaluation critique et outil d'évaluation critique
 - c) Étude descriptive : Dictionnaire de l'outil d'évaluation critique et outil d'évaluation critique
 - d) Revue de littérature : Dictionnaire de l'outil d'évaluation critique et outil d'évaluation critique
3. Exemple d'un tableau de synthèse des données probantes et de recommandations.

Consignes à l'intention des évaluateurs

Lors de l'étude des travaux de recherche, des questions sont formulées pour aider à déterminer quels documents sont nécessaires, à exposer des arguments et à présenter des recommandations. Aux fins de l'élaboration de lignes directrices, ces questions sont appelées **questions clés** (voir le glossaire). Avant de prendre la décision d'inclure ou non une étude, il faut la lire et déterminer brièvement ce qui a été fait. Si plus d'une question de recherche a été abordée ou que de multiples méthodes de recherche ont été utilisées, définissez les aspects qui se rapportent à votre question clé. Il faut noter qu'il est possible qu'un aspect d'une étude se rapporte à une question clé et qu'un autre aspect se rapporte à une différente question clé et que différentes méthodes de qualité variable soient employées. Une étude utilisée à l'appui de différentes conclusions doit être lue de nouveau en fonction de chaque question clé.

L'essentiel de l'évaluation critique est fondé sur l'interprétation des méthodes et des résultats, mais les sections d'analyse et de conclusions peuvent s'avérer utiles pour proposer d'autres explications des résultats, du biais, de la puissance, etc. Toutefois, en votre qualité d'évaluateur, les conclusions que vous tirez au sujet d'une étude doivent être fondées sur les méthodes et les résultats et non pas sur l'analyse et les conclusions de l'auteur.

Une fois qu'on a déterminé quels documents sont nécessaires, les études doivent faire l'objet d'une évaluation critique. La trousse d'outils a pour but d'aider à déterminer si les données probantes examinées suffisent à établir un lien entre l'exposition (p. ex. interventions, facteurs de risque, facteurs de protection ou démographiques) et le résultat en excluant d'autres explications pour le résultat signalé. Voici la marche à suivre :

1. Déterminez pourquoi vous passez l'article en revue. Lorsque vous lisez l'étude, prêtez attention aux méthodes et aux résultats qui se rapportent à votre question clé. Beaucoup d'études ont des résultats préliminaires et possiblement secondaires, et il est possible que seulement l'un de ces types de résultats vous intéresse, alors concentrez-vous sur celui-ci.
2. Lisez la section des méthodes dans l'étude pour obtenir un aperçu des méthodes de recherche employées. Si différents aspects d'une étude vous intéressent et que différentes méthodes ou différents plans d'étude ont été employés, alors vous devez évaluer chaque aspect séparément. Par exemple, une étude transversale peut servir à déterminer la prévalence, tandis qu'une étude cas-témoin nichée à une étude de cohorte peut servir à définir les facteurs de risque.
3. Nommez un plan d'étude en vous reportant aux méthodes utilisées pour déterminer le plan. Suivez chaque étape des algorithmes pour vous aider à sélectionner l'outil d'évaluation critique (OEC) approprié.
 - Nommer le plan et choisir l'OEC approprié vous aideront à vous assurer que vous évaluez les limites en fonction du plan particulier.
 - Pour les rapports d'éclosion, utilisez les algorithmes pour déterminer quel plan s'impose.
 - Si vous avez de la difficulté à nommer le plan d'étude, discutez-en avec des collègues et choisissez le plan qui vous permettra le mieux de définir les préoccupations les plus probables à évaluer.
 - N'acceptez pas le nom que l'auteur donne au plan de son étude à moins que vous ne soyez d'accord.
4. Décrivez le contenu de l'étude (en fonction de la question clé) dans le tableau de synthèse des données probantes.
 - Des lignes directrices pour identifier le contenu pertinent sont incluses.
 - Prêtez attention au contenu qui se rapporte à la question clé
5. Menez une évaluation critique de l'étude à l'aide du bon OEC.
 - Il y a trois types d'OEC, dont chacun est assorti d'un dictionnaire pour vous orienter dans le cadre du processus. Il importe de comprendre que ces dictionnaires ne présentent pas une explication en profondeur de tous les concepts et ne les illustrent pas à l'aide de tous les exemples possibles. Par conséquent, l'évaluateur doit utiliser son jugement pour interpréter les critères et les appliquer à l'étude faisant l'objet de l'examen et, en cas d'incertitude, en discuter avec ses collègues.
6. Ajoutez vos résultats et commentaires d'évaluation critique dans la dernière colonne du tableau de synthèse des données probantes.
7. Présentez une synthèse de la nature des études et des conclusions qui se rattachent à la question clé afin d'établir les recommandations. Les conclusions relatives à la qualité des données probantes sont généralement le fruit d'un consensus du groupe plutôt que d'une décision individuelle. Selon l'objet de votre évaluation critique, il peut être utile d'élaborer un compte rendu sommaire des données probantes et une justification pour le classement défini.

L'évaluation critique peut être chronophage, car elle exige que l'on prête attention aux détails et que l'on possède une expérience de l'évaluation critique de chaque élément. Les premières évaluations critiques que vous mènerez exigeront plus de temps, mais, avec l'expérience et la formation, vous pourrez mener des évaluations critiques plus rapidement. Les études analytiques prennent habituellement plus de temps à évaluer que les études descriptives, et il en va de même pour les articles complexes ou mal rédigés. Les discussions avec les collègues à différentes étapes sont utiles lorsqu'on mène une évaluation critique.

L'évaluation critique a pour objet de déterminer la qualité d'une étude. Cette trousse d'outils classe la qualité des études comme étant élevée, moyenne ou faible. La trousse d'outils offre assez de lignes directrices pour déterminer quelles questions doivent être analysées et laisse à l'évaluateur une marge de manœuvre pour appliquer les critères de l'évaluation critique. Il n'y a pas d'étude parfaite, et l'évaluation critique n'est pas une science exacte.

Tableau 1 – Définition des termes utilisés pour évaluer les données probantes

| | | |
|---|----------------------------|---|
| Force du plan d'étude Remarque : « x > y » signifie que le plan de x est plus fort que celui d'y. | Forte | Méta-analyse > essai contrôlé randomisé (ECR) > essai contrôlé non randomisé (ECNR) = expérience en laboratoire > étude comparatif avant-après (ECAA)* |
| | Modérée | Cohorte > cas-témoin> séries temporelles interrompues avec collecte de données adéquate > cohorte avec groupe de comparaison non équivalent |
| | Faible | Étude non-comparative avant-après (ENCAA) > séries temporelles interrompues avec collecte de données inadéquate > étude descriptive (transversale > sur le lien épidémiologique > écologique) |
| Qualité de l'étude | Élevée | Aucun grand obstacle à la validité interne (biais, hasard et confusion ayant fait l'objet d'un contrôle adéquat et été exclus à titre d'explication concurrente des résultats) |
| | Moyenne | Obstacles mineurs à la validité interne ne compromettant pas gravement la capacité de tirer une conclusion à l'égard de l'estimation de l'effet |
| | Faible | Grand(s) obstacle(s) à la validité interne qui compromettrait(aient) la capacité de tirer une conclusion à l'égard de l'estimation de l'effet |
| Nombre d'études | Multiples | Au moins 4 études |
| | Peu | Au plus 3 études |
| Uniformité des résultats | Uniformes | Études ayant abouti à des résultats semblables |
| | Disparates | Variation des résultats, mais tendance globale claire quant à l'effet |
| | Contradictoires | Résultats variables et aucune tendance globale claire quant à l'effet |
| Caractère direct des données probantes | Données probantes directes | Proviennent d'études portant précisément sur l'association étudiée |

| | | |
|--|---------------|--|
| | Extrapolation | Inférence tirée d'études portant sur une question clé différente, mais connexe, ou la même question clé dans des conditions artificielles (p. ex. certaines études en laboratoire) |
|--|---------------|--|

*La force du plan d'étude est forte s'il y a au moins deux groupes témoins et deux groupes d'intervention. La force est moyenne s'il y a seulement un groupe témoin et un groupe d'intervention.

Remarques

1. Certaines enquêtes sur **des éclosions ou sur des liens épidémiologiques** comprennent une comparaison/étude de groupes. Ces études sont par conséquent des études analytiques et devront y avoir une « force du plan » attribuée, puis les évaluer à l'aide de l'OEC d'une étude analytique. La majorité des études sur des éclosions et des études sur le lien épidémiologique ne contiennent pas de comparaison de groupes et sont donc des études descriptives.
2. **Les séries de cas et les rapports de cas** ne contribuent pas aux données probantes, par conséquent, on ne leur attribue pas de « force du plan » au moment de l'évaluation.
3. **Les études de modélisation** ne sont pas considérées dans ce classement, mais les évaluateurs doivent examiner la qualité des données sur lesquelles repose le modèle.

PARTIE 2 : OUTILS POUR NOMMER LE PLAN D'ÉTUDE, CLASSER LES DONNÉES PROBANTES ET RÉDIGER DES RECOMMANDATIONS

PLAN D'ÉTUDE

Le plan d'une étude désigne son « architecture », qui comprend des détails propres à la population étudiée, à l'échéancier, aux méthodes, aux procédures et aux considérations éthiques. La présente trousse d'outils n'aborde que les plans d'étude les plus couramment utilisés en recherche épidémiologique. Ces plans sont décrits dans les dictionnaires d'OEC respectifs, et certaines caractéristiques sont résumées dans le tableau 2 de la trousse d'outils. Tous les plans d'étude de la trousse d'outils appartiennent à l'une des trois catégories suivantes :

1. **Les études analytiques** sont conçues pour déterminer ou mesurer les effets d'une exposition, tels que des interventions ou des facteurs de risque. Le plan repose sur l'utilisation d'un groupe de comparaison approprié pour mettre à l'essai des hypothèses épidémiologiques, dans une tentative de définir des associations ou des relations causales.

Les études analytiques considérées comme interventionnelles ou expérimentales visent à évaluer les effets d'une intervention ou d'une mesure contrôlée par le chercheur; cela comprend les essais contrôlés randomisés (ECR), les essais contrôlés non randomisés (ECNR), les expériences en laboratoire et les études comparatives ou non comparatives avant-après (ECAA/ENCAA). Les études d'intervention comparent parfois deux ou plusieurs interventions. De telles études peuvent avoir un schéma d'expériences croisées, selon lequel les participants, après avoir subi une intervention, sont exposés à une autre intervention. Pour les besoins de la présente trousse d'outils, le schéma d'expériences croisées n'est pas considéré comme un plan d'étude en soi, mais pourrait plutôt s'inscrire dans la méthode d'un plan d'étude, comme un ECR, un ECNR ou un ECAA.

Les études analytiques considérées comme observationnelles sont des enquêtes scientifiques non expérimentales, qui reposent sur l'observation d'une situation, d'un comportement ou d'une intervention naturelle sans manipulation par le chercheur; il s'agit par exemple d'études de cohorte et d'études cas-témoin. Dans le cadre d'une étude de séries temporelles interrompues, le chercheur peut soit contrôler l'intervention, soit observer une situation.

2. **Les études descriptives** décrivent des caractéristiques générales ou particulières d'une affection en lien avec des facteurs particuliers ou l'exposition au phénomène étudié. Ces études portent sur la description, mais les chercheurs peuvent aussi mener une exploration préliminaire de l'association entre les variables; toutefois, ces études ne sont pas conçues pour éprouver des hypothèses. Ce plan permet souvent de découvrir les premiers indices importants se rattachant aux déterminants possibles d'une maladie et est surtout utile pour formuler des hypothèses qui peuvent ensuite être mises à l'épreuve à l'aide d'une étude analytique.
3. **Les revues de littérature** portent sur les points cruciaux d'un corpus de connaissances publiées. Cela s'accomplit au moyen de la synthèse, du classement et de la comparaison d'études analytiques passées ainsi que des revues de littérature et des articles théoriques. À l'exception des méta-analyses, qui consistent en une nouvelle analyse statistique de données provenant de plusieurs études mises en commun, ces études sont des sources secondaires et par conséquent ne contiennent pas de travail nouveau ou expérimental.

Chaque plan d'étude a ses forces et ses limites propres. Nommer le plan et choisir le bon OEC aideront à vous assurer que vous avez évalué les limites associées au plan particulier. Utilisez les outils pour nommer le plan d'étude afin de définir le plan et de sélectionner le bon OEC.

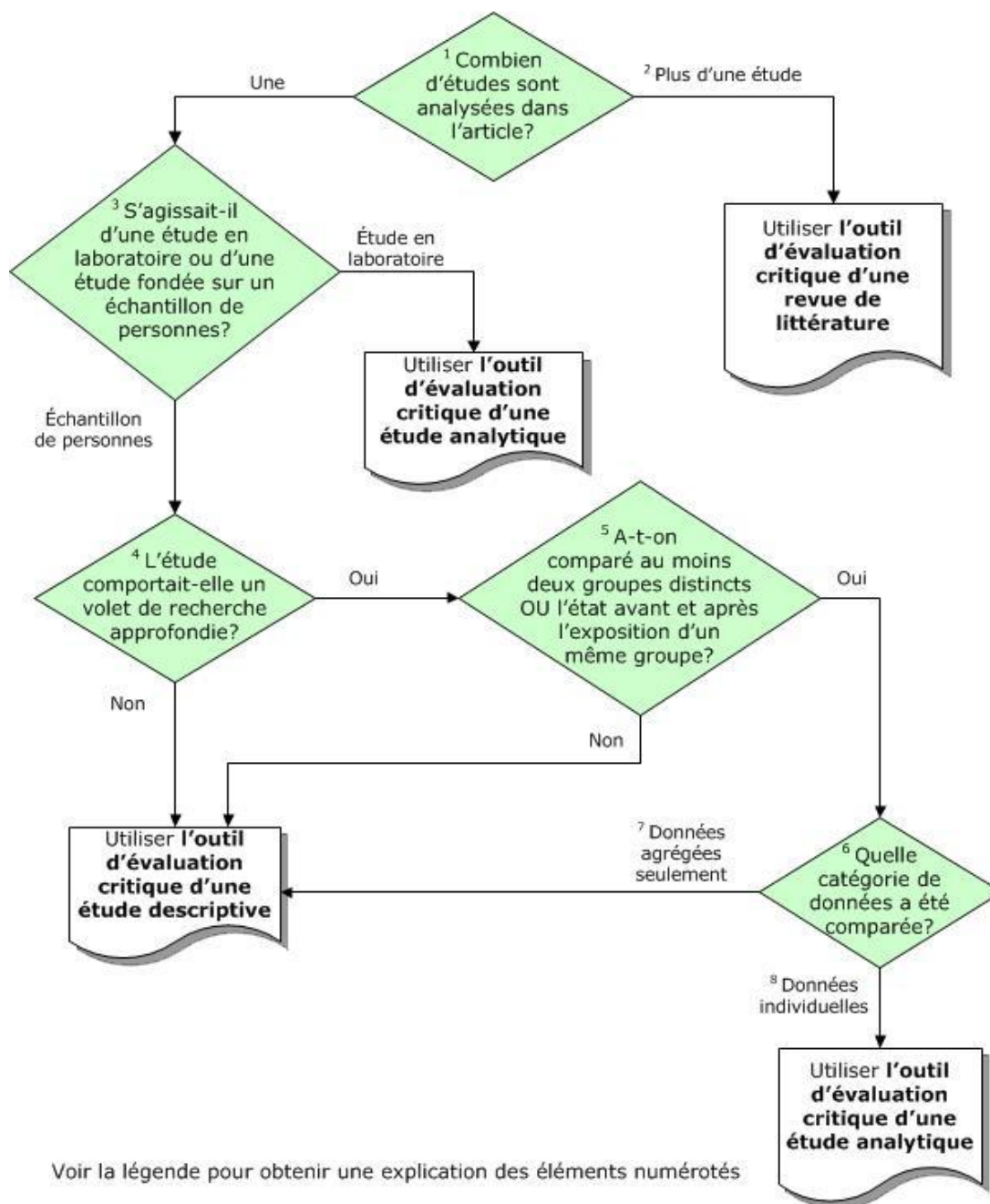
La terminologie employée pour nommer les plans d'étude en soins infirmiers et en sciences sociales peut varier, mais on peut habituellement appliquer les critères d'évaluation critique à la plupart des études même lorsqu'il est impossible de nommer précisément le plan. Dans cette trousse d'outils, le terme exposition se réfère aux expositions étudiées tel que les interventions, facteurs de risque, facteurs de protection ou démographiques, tandis que le résultat se réfère aux infections, maladies, comportements, effets ou affections.

Ces outils n'ont pas été conçus pour évaluer des études qui portent sur le rendement d'un outil diagnostique (p. ex. études qui évaluent la spécificité et la sensibilité). Il est possible d'appliquer les critères à de telles études, mais des outils plus appropriés peuvent être disponibles. Un autre type de plan d'étude qui n'est pas abordé dans la présente trousse d'outils est la **modélisation mathématique**. De telles études reposent sur une formule mathématique qui correspond à une équation et à des paramètres connexes servant à stimuler un processus, un système ou une relation. L'équation est créée à l'aide de sources de données primaires ou secondaires. En épidémiologie, les modèles mathématiques servent à expliquer ou à prévoir le résultat de la transmission d'une maladie, des interventions, des traitements ou des facteurs de risque.

Outils pour nommer le plan d'étude

La présente trousse d'outils contient **quatre algorithmes**; le premier algorithme aide à sélectionner le bon OEC, tandis que les trois autres aident à définir le plan d'étude. Une légende des éléments numérotés est fournie avec chaque algorithme.

Figure 1 : Algorithme pour sélectionner le bon outil



Légende : Algorithme pour sélectionner le bon outil

1. **Combien d'études sont analysées dans l'article?** L'évaluation critique d'une étude unique diffère de l'évaluation critique d'un corpus de données probantes (plus d'une étude). Différents outils sont à votre disposition pour vous orienter dans le processus d'évaluation critique. Pour choisir le bon outil, il faut tout d'abord déterminer si l'article que vous lisez porte sur une seule étude ou plusieurs.
2. **Plus d'une étude.** Les revues de littérature, les synthèses de littérature, les revues systématiques de littérature, les méta-analyses et les lignes directrices sont des exemples d'articles qui portent sur plusieurs études à la fois. Ils devraient être évalués à l'aide de l'OEC d'une revue de littérature. Utilisez « l'algorithme pour nommer le type de revue de littérature » pour définir le type de revue de littérature.
3. **S'agissait-il d'une étude en laboratoire ou d'une étude fondée sur un échantillon de personnes?** Les études en laboratoire sont généralement des expériences contrôlées et entraînent donc l'utilisation de l'OEC d'une étude analytique. L'emploi d'une technique ou d'un essai en laboratoire pour obtenir de l'information ne signifie pas qu'il s'agit d'une étude en laboratoire; par exemple, des études qui contiennent des résultats obtenus en laboratoire comme le typage moléculaire dans le contexte d'une infection ou de la contamination de patients à la suite d'une éclosion ne sont pas des études en laboratoire. Inversement, les expériences en laboratoire peuvent supposer la participation d'un échantillon de personnes dans un laboratoire.
4. **L'étude comportait-elle un volet de recherche approfondie?** L'étude peut se limiter à une description de l'incidence ou comprendre une enquête sur un lien, p. ex., entre les cas ou les affections.
5. **A-t-on comparé au moins deux groupes distincts ou l'état avant et après l'exposition d'un même groupe?** Dans le cadre d'une étude analytique, on vérifie une hypothèse quant aux effets d'une exposition dans un groupe comparativement à un groupe témoin. Une étude peut évaluer l'exposition étudiée (p. ex. facteurs de risque, interventions, facteurs de protection ou démographiques) et/ou ses résultats (p. ex. infections, maladies, comportements, effets ou affections) dans plus d'un groupe cible.
 - Dans le cadre d'une étude d'intervention (comme un essai contrôlé), lorsqu'au moins deux groupes font l'objet d'une comparaison dans une étude sur les effets d'une exposition, on parle habituellement du groupe témoin ou de comparaison et du groupe d'intervention ou expérimental. Différents termes sont employés dans une étude d'observation (p. ex. cas-témoin ou cohorte) pour évaluer les facteurs de risque ou de protection ou les interventions naturelles non manipulées par le chercheur. Dans le cadre d'une étude cas-témoin, il y a un groupe cas et un groupe témoin. Dans une étude de cohorte, on parle du groupe exposé et du groupe non exposé.
 - Dans certaines études, un groupe peut être évalué avant l'exposition et de nouveau après l'exposition, ou le groupe avant l'exposition et le groupe après l'exposition peuvent être composés de différentes personnes. Dans d'autres études, il peut y avoir plusieurs périodes d'évaluation après l'exposition. Même lorsqu'il y a seulement un groupe évalué à différents moments plutôt que deux groupes simultanément, on doit considérer l'étude comme établissant une comparaison entre des groupes.

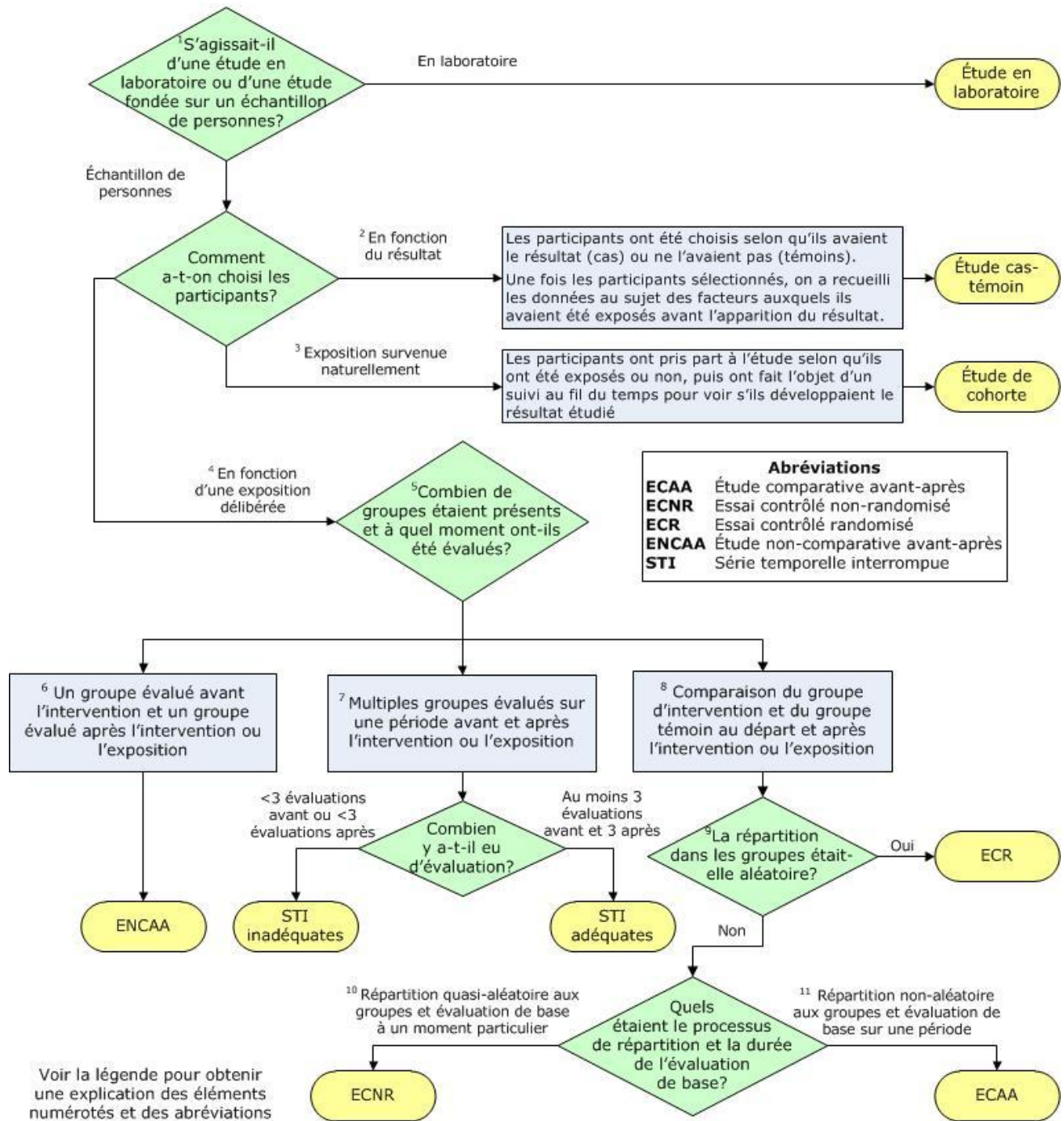
- Le type de plan employé varie dans les **enquêtes sur des éclosions ou le lien épidémiologique**. On ne peut pas automatiquement classer les études sur le lien épidémiologique dans la catégorie des études descriptives ou analytiques; il faut les lire attentivement. Le type de plan peut seulement être attribué au cas par cas. Lorsqu'il y a une comparaison de groupes dans une enquête sur des éclosions (comme dans le cadre d'une étude de cohorte ou une étude cas-témoin), on considère qu'il s'agit d'une étude analytique. Une enquête sur des éclosions sans comparaison de groupes est considérée comme une étude descriptive.

En général, les études descriptives n'ont pas de groupe témoin, bien que l'analyse menée dans le cadre d'une étude transversale puisse comprendre une comparaison des résultats chez des personnes qui respectent des facteurs particuliers associés à l'étude. Si seulement un groupe est évalué à tout moment et qu'il n'y a pas de groupe témoin, il faut utiliser « l'algorithme pour nommer le type d'étude descriptive » afin de définir le plan d'étude, puis mener l'évaluation à l'aide de l'outil d'évaluation critique d'une étude descriptive.

6. **Quelle catégorie de données a été comparée?** Il faut tenir compte de la catégorie de données lorsqu'on compare deux groupes distincts ou le même groupe avant et après un événement. Pour comprendre le risque à l'égard d'une personne, il faut déterminer si le résultat étudié est survenu chez les personnes exposées au facteur de risque ou à l'intervention étudié. Par conséquent, il faut faire la distinction entre les données individuelles et les données agrégées.
7. **Données agrégées seulement.** Dans le cadre d'une étude **écologique**, seules les données agrégées peuvent être analysées, et non pas les données individuelles. Il est impossible d'associer un résultat à une exposition chez un sujet particulier. Par exemple, les résultats de la surveillance doivent être disponibles pour déterminer le nombre de cas de grippe dans une région, et des données relatives à la vaccination peuvent être disponibles pour décrire la couverture de vaccination contre la grippe offerte dans la région particulière. Dans le cadre d'une étude écologique, il est possible d'analyser les taux dans un groupe, mais il n'y a pas de données disponibles permettant de déterminer si les individus étaient exposés ou non (p. ex. vaccin contre la grippe) et ont obtenu le résultat ou non (p. ex. la grippe).
8. **Données individuelles.** Une étude qui compare l'exposition et le résultat chez des sujets appartenant à deux groupes différents sert à vérifier des hypothèses d'association entre l'exposition et le résultat. Il s'agit d'études analytiques. Il faut employer « l'algorithme pour nommer le type d'étude analytique » pour définir le plan de l'étude, puis mener l'évaluation à l'aide de l'outil d'évaluation critique d'une étude analytique.

Les études en laboratoire sont considérées comme des études analytiques.

Figure 2 : Algorithme pour nommer le type d'étude analytique



Légende : Algorithme pour nommer le type d'étude analytique

1. **S'agissait-il d'une étude en laboratoire ou d'une étude fondée sur un échantillon de personnes?** Les études en laboratoire sont généralement des expériences contrôlées, comparant les résultats d'au moins deux ensembles différents d'affections. L'emploi d'une technique ou d'un essai en laboratoire pour obtenir de l'information ne signifie pas qu'il s'agit d'une étude en laboratoire; par exemple, des études qui contiennent des résultats obtenus en laboratoire comme le typage moléculaire dans le contexte d'une infection ou de la contamination de patients à la suite d'une éclosion ne sont pas des études en laboratoire. Inversement, les expériences en laboratoire peuvent supposer la participation d'un échantillon de personnes dans un laboratoire ou un milieu artificiel.
2. **Comment a-t-on choisi les participants? En fonction du résultat.** La grande distinction entre les études cas-témoin et les autres types d'études analytiques tient au fait que, dans une étude cas-témoins, les participants sont sélectionnés en fonction du résultat plutôt que de l'exposition. Par exemple, on peut trouver des patients avec (cas) et sans (témoin) grippe clinique puis les comparer en fonction de l'exposition préalable au vaccin contre la grippe. Les cas peuvent provenir d'une cohorte (étude cas-témoins niché). Avez votre analyse sur les cas-témoins ou la cohorte selon le résultat étudié.
3. **Comment a-t-on choisi les participants? En fonction de l'exposition survenue naturellement.** Dans le cadre de presque toutes les études analytiques (à l'exception des études cas-témoins), les participants ont été suivis à partir de l'exposition pour déterminer s'ils atteignaient le résultat étudié. Une autre distinction entre les plans d'étude s'établit selon que l'exposition est survenue naturellement ou non.
4. **Comment a-t-on choisi les participants?** En fonction d'une exposition délibérée. L'exposition n'est pas survenue naturellement; elle a plutôt été déterminée par le chercheur.

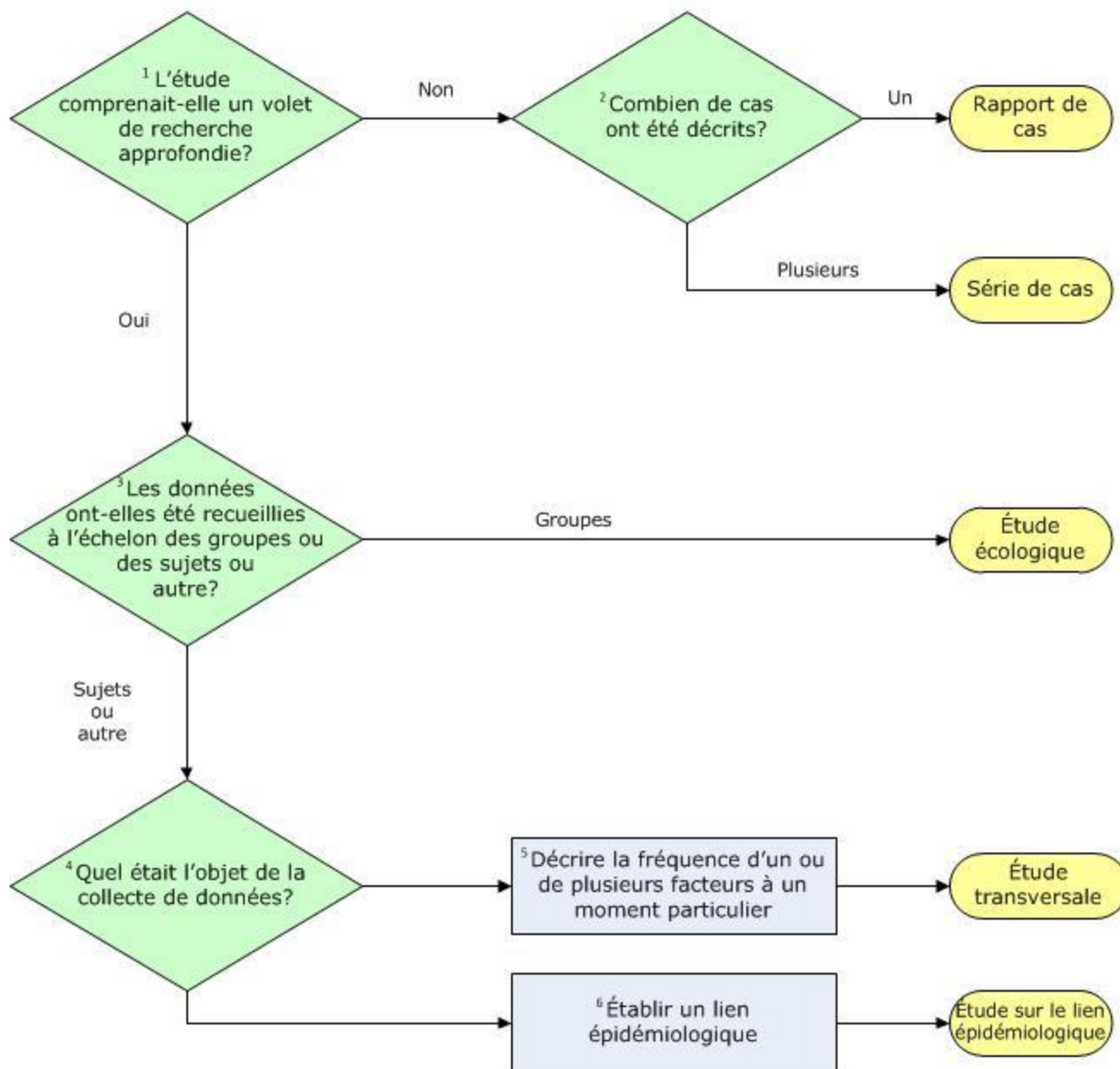
Il importe de noter que, peu importe la méthode de sélection des participants – en personne ou à partir d'une base de données, comme une base de données de surveillance, un registre ou autre, elle n'a aucune incidence sur le nom du plan d'étude. Ce fait demeure vrai que les participants aient été choisis en fonction du résultat, de l'exposition naturelle ou à la suite d'une exposition délibérée.

5. **Combien de groupes étaient présents et à quel moment ont-ils été évalués?** Il importe de déterminer s'il y avait plus d'un groupe et plus d'une période d'évaluation afin de nommer le plan.
6. **Un groupe évalué avant l'intervention et un groupe évalué après l'intervention.** Dans le cadre d'une étude non comparative avant-après (ENCAA), il n'y a pas de groupe témoin simultané. Il y a eu une intervention auprès d'un groupe de participants, et les résultats avant et après l'intervention sont comparés. Les sujets du groupe après l'intervention ne sont pas nécessairement les mêmes que ceux du groupe avant l'intervention. Si les sujets sont les mêmes, les données peuvent être comparées pour chacun avant et après l'intervention. Cela améliore la qualité des données probantes, mais le plan est toujours considéré comme étant faible grâce au groupe témoin inadéquat.

7. **Multiplés groupes évalués sur une période avant et après l'intervention ou l'exposition** (séries temporelles interrompues ou STI) : Méthode communément associée à la surveillance. Pour établir des statistiques sur les tendances appropriées et pour que le travail soit considéré comme des STI adéquates, il est essentiel d'avoir au moins trois points de données d'évaluation de base et trois points de données après l'intervention. Il importe aussi de pouvoir définir clairement à quel moment l'exposition (p. ex. intervention, facteur de risque ou autre) est survenue.
8. **Comparaison du groupe d'intervention et du groupe témoin au départ et après l'intervention.** Les études d'intervention se distinguent en outre par le fait que les participants ont été répartis aléatoirement ou non au groupe d'intervention et au groupe témoin et la nature de l'évaluation avant l'intervention. Voir les descriptions 9, 10 et 11 ci-bas.
9. **La répartition dans les groupes était-elle aléatoire? Oui.** Dans le cadre d'un essai contrôlé randomisé (ECR), le chercheur répartit les participants aux groupes de façon aléatoire, p. ex. par la génération de chiffres aléatoires ou en tirant à pile ou face. La répartition aléatoire permet de mieux contrôler les variables confusionnelles inconnues. Si les auteurs affirment avoir réparti les sujets aux groupes de façon aléatoire, il faut parler d'un ECR et évaluer la qualité de la répartition aléatoire en fonction des critères établis au numéro 9 de l'outil d'évaluation critique d'une étude analytique.
10. **La répartition quasi aléatoire aux groupes :** Dans le cadre d'un essai contrôlé non randomisé (ECNR), les participants sont assignés au groupe d'intervention et au groupe témoin de façon systématique qui ne reflète pas réellement la répartition aléatoire, p. ex. alternance entre les groupes ou utilisation de l'année de naissance. L'évaluation de base se fait à un moment particulier.
11. **Répartition non aléatoire aux groupes et évaluation de base sur une période :** Dans le cadre d'une étude comparative avant-après (ECAA), il n'y a pas de répartition aléatoire ou quasi aléatoire aux groupes. En général, les participants sont répartis de façon naturelle, p. ex. ils travaillent ensemble dans la même région. Une ECAA avec deux groupes témoins et deux groupes d'intervention a un meilleur contrôle du biais potentiel comparé à une ECAA avec un groupe témoin et un groupe d'intervention. Il faut tenir compte du nombre de groupes afin de distinguer une étude comparative avant-après (nombre limité de groupes) d'un essai randomisé par grappe. Dans ce cas, la répartition aléatoire se fait à l'échelon du sous-groupe (p. ex. par unité de soins), mais il y a de nombreux sous-groupes dans le groupe d'intervention et le groupe témoin. Dans le cadre d'une ECAA, il y a aussi une période d'évaluation de base; l'évaluation de base ne survient donc pas à un moment particulier.

Si, après avoir suivi l'algorithme, vous n'êtes toujours pas certain du plan d'étude, vous pouvez consulter le tableau 2 ou discuter de l'étude avec vos collègues.

Figure 3 : Algorithme pour nommer le type d'étude descriptive



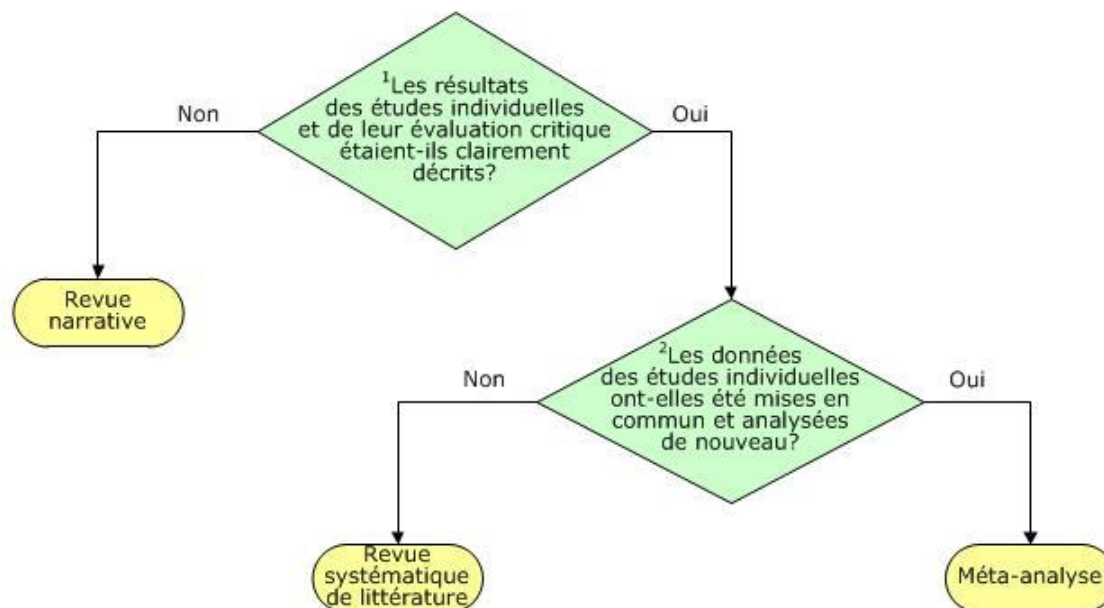
Voir la légende pour obtenir une explication des éléments numérotés

Légende : Algorithme pour nommer le type d'étude descriptive

Remarque : Si le plan d'étude n'est pas clair (mais qu'il ne s'agit manifestement pas d'une étude analytique), il faut parler d'une étude descriptive et utiliser l'outil d'évaluation critique d'une étude descriptive.

1. ***L'étude comportait-elle un volet de recherche approfondie?*** L'étude peut se limiter à une description de l'incidence ou comprendre une enquête sur un lien, p. ex., entre les cas ou les affections. Le volet de recherche approfondie offre une dimension additionnelle aux études descriptives.
2. ***Combien de cas ont été décrits?*** Un rapport de cas est une description détaillée de l'expérience d'un patient individuel, tandis qu'une série de cas décrit l'expérience de plusieurs patients souffrant de la même maladie, étant exposés au même facteur ou possédant les mêmes caractéristiques.
3. ***Les données ont-elles été recueillies à l'échelon des groupes ou des sujets?*** Il faut tenir compte de la catégorie de données si on a établi une comparaison entre deux groupes distincts ou auprès d'un même groupe avant et après un évènement. Pour comprendre le risque chez un sujet, il faut déterminer si le résultat étudié est survenu chez le sujet exposé au facteur de risque ou à l'intervention étudié. Ainsi, il faut établir la distinction entre les données individuelles et les données agrégées. Les données peuvent être des microbes, des facteurs environnementaux, des gens, etc. Certaines études, les études écologiques, comparent les données de l'exposition et du résultat à l'échelon de la population ou à l'échelon agrégé, p. ex. on peut vérifier si des isolats de microorganismes résistant aux antibiotiques en laboratoire augmentent en fonction de la consommation d'antibiotiques à l'hôpital. Toutefois, on ne sait pas si le résultat est survenu chez les sujets exposés au facteur étudié. On obtient des données individuelles dans des études qui comparent l'exposition et le résultat chez des sujets appartenant à deux groupes différents, puis on les utilise ensuite pour vérifier des hypothèses quant aux associations entre l'exposition et le résultat.
4. ***Quel était l'objet de la collecte de données?*** La collecte de données peut avoir une ou plusieurs raisons d'être, comme la description des facteurs préexistants ou l'établissement d'un lien épidémiologique.
5. ***Décrire la fréquence d'un ou de plusieurs facteurs à un moment particulier.***
 - Une étude transversale décrit l'exposition et le résultat étudiés chez les sujets d'un groupe au même moment; elle présente un instantané ou un profil à un moment défini. Par exemple, dans le cadre d'un sondage, on pourrait demander à quelqu'un s'il a eu des symptômes de rhume (résultat) dans la semaine de l'étude et s'il a pris de la vitamine C (exposition) durant la même semaine. Une étude transversale ne permet pas toujours de déterminer si l'exposition a précédé le développement du résultat.
 - Les études transversales qui décrivent le nombre de sujets exposés (p. ex. tabagisme) ou présentant le résultat (p. ex. infection) étudié au moment de la collecte de données sont parfois désignées d'études de prévalence. La prévalence quantifie la proportion de personnes dans une population qui ont le résultat ou qui ont été exposées au facteur étudié à un moment donné ou durant une période particulière.

- Il faut bien s'assurer d'établir la distinction entre l'étude transversale et l'étude de cohorte rétrospective. Dans toutes les études de cohorte, qu'elles soient rétrospectives ou prospectives, les résultats sont comparés pour les personnes qui ont été exposées au facteur étudié dès le début de l'étude et celles qui n'ont pas été exposées. Il n'y a pas de résultat de base pour les participants des groupes exposé et non exposé et l'exposition est survenue naturellement (c.-à-d., que le chercheur n'a pas contrôlé l'exposition). Une étude de cohorte rétrospective s'appuie sur un ensemble de données existant (recueilli par le passé) et assure le suivi des participants à compter d'un point de départ jusqu'à un moment donné déterminés au préalable (qui peuvent aussi être des moments passés, p. ex., suivi des participants de 2002 à 2005). Une étude de cohorte prospective recueille des données à compter du moment actuel (début de l'étude) jusqu'à un moment donné.
6. **Établit un lien épidémiologique.** Aux fins de la présente trousse d'outils, les études sur le lien épidémiologique sont une catégorie d'études descriptives composée de recherches rétrospectives de la source, des patients exposés et les contacts. Les sujets de ces études font l'objet d'une évaluation pour trouver des liens (p. ex. contact ou typage microbien) vers des cas, des contacts ou des conditions.
- Pour les **enquêtes sur des éclosions**, le plan d'étude ne peut être assigné qu'individuellement. La plupart des études sur des éclosions ne comportent pas de groupe témoin; celles qui en comptent un sont considérées comme étant des études analytiques donc il faut se référer à la figure 2 pour nommer le type de plan d'étude.

Figure 4 : Algorithme pour nommer le type de revue de littérature**Légende : Algorithme pour nommer le type de revue de littérature**

1. **Les résultats des études individuelles et de leur évaluation critique étaient-ils clairement décrits?** Une **revue narrative** n'offre pas d'information sur l'évaluation critique des études individuelles de l'examen, mais résume simplement ces études et en interprète les résultats. Une **revue systématique de littérature** offre des détails sur le processus d'évaluation critique et les résultats liés aux études individuelles.
2. **Les données des études individuelles ont-elles été mises en commun et analysées de nouveau?** La différence entre une revue systématique de littérature et une **méta-analyse** tient au fait que, dans le cadre d'une méta-analyse, on met en commun de façon quantitative les données des études individuelles visées par l'examen systématique de littérature.

Tableau 2 – Plans d'étude analytique et descriptive

Plans d'étude présentés en ordre décroissant selon la force du plan
(comme indiqué dans le tableau 1)

| Plan d'étude et outil | Admission à l'étude, groupes | Évaluation de base effectuée | Exposition* | Comparaison des résultats |
|--|---|--|---------------------------|---|
| Essai contrôlé randomisé (ECR) <i>Analytique</i> | <ul style="list-style-type: none"> Recrutement avant l'intervention Affectation au groupe témoin ou au groupe d'intervention (expérimental) en fonction d'un processus aléatoire (p. ex. génération de chiffres aléatoires, tirer à pile ou face) | Habituellement mesurée à un moment donné au moment de l'admission dans l'étude pour évaluer la présence préalable du résultat et d'autres caractéristiques chez le sujet | Contrôle par le chercheur | Chaque personne est suivie et évaluée en fonction des résultats, et les résultats sont comparés entre les groupes. |
| Essai contrôlé non randomisé (ECNR) <i>Analytique</i> | <ul style="list-style-type: none"> Recrutement avant l'intervention Affectation au groupe témoin ou au groupe d'intervention Répartition selon un processus quasi aléatoire (p. ex. alter-nance) | Habituellement mesurée à un moment donné au moment de l'admission dans l'étude pour évaluer la présence préalable du résultat et d'autres caractéristiques chez le sujet | Contrôle par le chercheur | Chaque personne est suivie et évaluée en fonction des résultats, et les résultats sont comparés entre les groupes. |
| Expérience en laboratoire <i>Analytique</i> | <ul style="list-style-type: none"> Situation artificielle Groupes témoins et expérimentaux ou conditions en laboratoire | Oui | Contrôle par le chercheur | Les résultats sont évalués et comparés entre les groupes témoins ou expérimentaux ou les conditions en laboratoire. |
| Étude comparative avant-après (ECAA) <i>Analytique</i> | <ul style="list-style-type: none"> Recrutement avant l'intervention Affectation à un groupe témoin ou à un groupe d'intervention Affectation selon un processus non aléatoire (p. ex. groupes naturels, comme les employés d'une unité de soins) | Oui, il y a une période de base pour l'évaluation (pas seulement à un moment donné à l'admission) | Contrôle par le chercheur | Chaque personne est suivie et évaluée en fonction des résultats, et les résultats sont comparés entre les groupes. |

| Plan d'étude et outil | Admission à l'étude, groupes | Évaluation de base effectuée | Exposition* | Comparaison des résultats |
|--|---|---|--|---|
| Cohorte <i>Analytique</i> | <ul style="list-style-type: none"> L'admission dépend de l'exposition préalable On parle du groupe exposé et non exposé | Oui, les participants sont réputés ne pas avoir le résultat au début de l'étude | L'exposition est survenue naturellement (pas contrôlée par le chercheur). Suivi pour déterminer si le résultat survient. | Chaque personne est suivie et évaluée en fonction des résultats, et les résultats sont comparés entre les groupes exposés et non exposés. |
| Cas-témoin <i>Analytique</i> | <ul style="list-style-type: none"> L'admission dépend de l'obtention du résultat au préalable Les groupes sont des cas s'ils ont les résultats et témoins dans le cas contraire Les cas sont souvent tirés d'une cohorte (étude de cas-témoins niché), mais la sélection des cas et des témoins est fondée sur le résultat plutôt que l'exposition | Aucune | Les expositions étudiées et les résultats sont survenus naturellement (pas contrôlés par le chercheur). | Chaque personne est évaluée en fonction de l'exposition, et on compare les résultats des cas et des témoins. |
| Série temporelle interrompue (STI) <i>Analytique</i> | <ul style="list-style-type: none"> Différents groupes à différents moments, p. ex. surveillance Chaque groupe est probablement composé de différents sujets | Trois points de données distincts avant l'intervention sont nécessaires pour une analyse adéquate des tendances | L'exposition peut être un phénomène naturel ou contrôlé par le chercheur. Il doit y avoir un moment clairement défini où est survenue l'exposition ou l'intervention. | L'exposition ou le résultat est mesuré auprès du sujet individuel, mais il peut être déclaré en tant qu'agrégat. Les résultats sont comparés entre les périodes. Trois points de données distincts après l'intervention sont aussi nécessaires. |

| Plan d'étude et outil | Admission à l'étude, groupes | Évaluation de base effectuée | Exposition* | Comparaison des résultats |
|---|---|--|--|--|
| Étude non comparative avant-après (ENCAA) <i>Analytique</i> | <ul style="list-style-type: none"> Il y a seulement un groupe à la fois. Le groupe subit une évaluation à la base, est exposé à l'intervention et est évalué de nouveau par la suite Les groupes avant et après ne sont pas nécessairement composés des mêmes sujets. Il n'y a pas de groupe témoin simultané, bien que les sujets fassent office de témoins correspondant au bout du compte | Oui | L'exposition peut être un phénomène naturel, mais est habituellement contrôlée par le chercheur. | Les résultats sont comparés en fonction des deux périodes. |
| Étude transversale <i>Descriptive</i> | <ul style="list-style-type: none"> Un groupe à un moment donné | Les données sont recueillies quant à l'exposition et au résultat au même moment | L'exposition et le résultat surviennent naturellement (pas contrôlés par le chercheur). | On établit une comparaison entre ceux chez qui on observe le résultat et ceux chez qui on ne les observe pas. |
| Étude sur le lien épidémiologique <i>Descriptive</i> | <ul style="list-style-type: none"> Admission en fonction du contact connu ou soupçonné avec une source ou un sujet infecté | L'évaluation de base ne fait pas partie de l'étude, mais elle peut avoir été menée dans le cadre de pratiques cliniques. | Exposition survenue naturellement. | Les sujets sont évalués pour déterminer s'il existe des liens (p. ex. contact ou typage microbien) vers des cas, des contacts ou des conditions. |
| Étude écologique (ou corrélative) <i>Descriptive</i> | <ul style="list-style-type: none"> Différents groupes à différents moments | Oui | Exposition survenant habituellement naturellement | L'exposition et le résultat sont mesurés de façon agrégée. |

| Plan d'étude et outil | Admission à l'étude, groupes | Évaluation de base effectuée | Exposition* | Comparaison des résultats |
|---|--|------------------------------|---|--|
| Rapport de cas /série de cas <i>Descriptive</i> | <ul style="list-style-type: none"> Description des expériences d'un sujet (de cas) ou de quelques sujets (séries de cas). Admission fondée sur le résultat | Oui ou non | Exposition et résultat survenant naturellement (pas contrôlés par le chercheur) | Aucune comparaison de groupes. On fournit une description des cas chez lesquels on a observé le résultat. Il faut mener des travaux de recherche plus en profondeur pour déterminer s'il y a une association entre les résultats et les expositions possibles. |

*Voir le glossaire pour la définition

Tableau de synthèse des données probantes et rédaction de recommandations

Un tableau de synthèse des données probantes simplifie l'examen des études en tant que corpus de données probantes en fonction d'un argument se rattachant à la question clé. Au premier coup d'œil, il est possible d'effectuer une comparaison entre les études, les plans, les tailles d'échantillon, les interventions, les similitudes des participants et les mesures de résultats. On obtient également les enjeux pertinents en ce qui concerne les forces et les limites de l'étude, la synthèse des résultats en fonction de la question clé et les conclusions. Le tableau de synthèse des données probantes contribue à l'analyse de la force des données probantes, des enjeux définis par l'intermédiaire de l'évaluation critique et des recommandations.

Dans le cadre du processus d'élaboration des lignes directrices, on a besoin de deux personnes pour mener une évaluation critique de chaque étude. Les évaluateurs affectés à une question clé donnée et aux études qui s'y rapportent sont responsables de remplir le tableau de synthèse des données probantes à l'aide des données liées à la question clé.

Les rangées d'un tableau de synthèse des données probantes contiennent de l'information sur les études individuelles, tandis que les colonnes reflètent le type de détails à inclure. Il importe de noter que la deuxième colonne du tableau de synthèse des données probantes qui figure à l'annexe B, intitulée « Méthodes et mesures des résultats pertinentes », peut être divisée en deux, selon la quantité de données à inclure. L'utilisation d'un format « paysage », de puces (plutôt que des phrases complètes) et d'abréviations acceptables contribue à l'efficacité. Le tableau 3 présente une synthèse du contenu qui devrait être ajouté au tableau de synthèse des données probantes.

Il faut noter que les études devraient être énumérées en ordre décroissant dans le tableau selon la force du plan, les méta-analyses étant en tête de liste et les études de cas étant mentionnées en dernier, voire exclues. Le corpus de données probantes complet (pas seulement une étude) devrait faire l'objet d'un examen et d'une discussion avec vos collègues avant la formulation d'une recommandation.

Tableau 3 – Contenu pertinent pour un tableau de synthèse des données probantes

Question clé : _____

| Auteur (année) Numéro de référence | Méthodes et mesures des résultats pertinentes | Résultats | Conclusions* Commentaires de l'évaluateur Classement de l'étude |
|--|--|--|---|
| | <ul style="list-style-type: none"> • Pays • Contexte (p. ex. unité de soins intensifs, soins pédiatriques, milieu rural ou urbain) • Nombre de participants dans le groupe témoin et dans le groupe d'intervention ou taille globale de l'échantillon en l'absence de groupes distincts • Caractéristiques particulières ou principales de l'échantillon • Détails particuliers de l'exposition (p. ex. interventions, facteurs de risque, facteurs de protection ou démographiques) se rapportant aux questions clés • Méthodes de collecte de données et moment de la mesure • Mesures utilisées • Validité, fiabilité ou fidélité interévaluateurs le cas échéant | <ul style="list-style-type: none"> • Résultats qui se rapportent à la question clé (résultat étudié primaire et/ou secondaire) • Résultats particuliers (p. ex. proportion ou moyenne), intervalle de confiance et valeur <i>P</i>, le cas échéant | <ul style="list-style-type: none"> • Conclusions globales de l'étude dont la force du plan, la qualité de l'étude et le caractère direct des données probantes • Principales forces qui auraient une incidence sur les décisions • Principales limites qui auraient une incidence sur l'interprétation des résultats • Toute préoccupation que l'évaluateur souhaite noter ou dont il veut discuter avec le groupe de travail chargé d'élaborer des lignes directrices • Commentaires sur l'application de l'intervention ou des résultats à des fins de généralisation à d'autres groupes et de détermination de la faisabilité de la mise en œuvre |

* Les conclusions liées à la force du plan, à la qualité de l'étude et au caractère direct des données probantes devraient être fondées sur les définitions des termes utilisés pour évaluer les données probantes (tableau 1).

Remarque : Les dictionnaires relatifs aux différents outils d'évaluation critique offrent une certaine orientation quant aux éléments à inscrire dans la colonne des conclusions et des commentaires. Le corpus de données probantes, et si nécessaire, des études individuelles complexes, devraient faire l'objet d'un examen en consultation avec des collègues.

Formuler une recommandation

Le tableau de synthèse des données probantes permet aux évaluateurs de constater l'ampleur et l'uniformité d'un effet d'une étude à l'autre et de tirer une conclusion. La prochaine étape varie en fonction de l'objectif de l'évaluation critique. Par exemple, si on élabore des lignes directrices, l'objectif sera de présenter une recommandation visant la pratique et d'y attribuer des données probantes d'un certain classement. Le tableau 4 résume les critères de classement des données probantes pour cette trousse d'outils.

Lorsqu'ils font des recommandations, les évaluateurs devraient s'assurer qu'elles sont

- fondées sur une conclusion valide tirée des données probantes accessibles;
- formulées à la voix active.
- Voici d'autres aspects à considérer
- quantité, qualité et uniformité des données probantes;
- incidence de la recommandation sur la pratique et coût si elle est mise en œuvre;
- faisabilité de la mise en œuvre

Remarque concernant le classement des données probantes

Lorsqu'une recommandation est fondée sur un règlement, il n'y a pas lieu d'y attribuer un classement.

Les classements se rattachent aux données probantes et non pas à la recommandation.

Le classement attribué à une revue systématique de littérature varie en fonction de l'évaluation critique de tous les aspects des données probantes rapportées (force du plan d'étude, qualité de l'étude, nombre d'études visées par la revue, uniformité des résultats d'une étude à l'autre et caractère direct des données probantes). Si l'évaluation critique démontre que la revue systématique de littérature était de qualité élevée, il faut intégrer les résultats aux données probantes et leur attribuer un classement. Si la revue systématique de littérature se révèle de qualité moyenne ou faible, il faut bien réfléchir avant de déterminer si les résultats seront intégrés aux données probantes.

Synthèse des données probantes dans un texte

Selon l'objet de l'évaluation critique, il peut être utile de présenter une synthèse des données probantes et une justification du classement attribué. Cela devrait comprendre la réponse à la question clé et des données pertinentes liées aux éléments suivants

- nombre d'études incluses dans la synthèse des données probantes;
- types d'études incluses (p. ex. nombre d'études par plan et force);

- relevé sommaire des résultats (p. ex. uniformes ou variables, tendances ou signification statistique, effet ou absence d'effet);
- force des données probantes globales à l'appui d'une recommandation;
- enjeux à considérer avant de formuler une recommandation (p. ex. si on devrait obtenir plus de littérature, faisabilité, coûts, etc.).

Tableau 4 – Critères pour classer les données probantes à la source des recommandations

| Classement des données probantes | | |
|----------------------------------|------------|---|
| Force des données probantes | Classement | Critères |
| Forte | AI | Données probantes directes provenant d'une méta-analyse ou de multiples études à modèle fort de haute qualité, avec résultats cohérents |
| | AII | Données probantes directes provenant de multiples études à modèle fort de qualité moyenne, avec résultats cohérents ou Au moins une étude à modèle fort étayée par de multiples études à modèle modéré de haute qualité, avec résultats cohérents ou Au moins une étude à modèle fort de qualité moyenne étayée par une extrapolation de multiples études à modèle fort de haute qualité, avec résultats cohérents |
| Modérée | BI | Données probantes directes provenant de multiples études à modèle modéré de haute qualité, avec résultats cohérents ou Extrapolation de multiples études à modèle fort de haute qualité, avec résultats cohérents |
| | BII | Données probantes directes provenant d'un ensemble d'études à modèle fort ou modéré de qualité élevée ou moyenne, avec une tendance claire, mais certains résultats incohérents ou Extrapolation de multiples études à modèle fort de qualité moyenne ou d'études à modèle modéré de qualité élevée ou moyenne, avec résultats cohérents ou Une étude à modèle fort étayée par de multiples études à modèle faible de qualité élevée ou moyenne, avec résultats cohérents |
| Faible | CI | Données probantes directes provenant de multiples études à modèle faible de qualité élevée ou moyenne, avec résultats cohérents ou Extrapolation de l'association d'études à modèle fort ou modéré de qualité élevée ou moyenne, avec résultats incohérents |
| | CII | Étude de faible qualité, quel que soit le modèle ou Résultats contradictoires, quel que soit le modèle ou Séries de cas ou rapports de cas ou Opinion d'expert |

PARTIE 3 : OUTILS D'ÉVALUATION CRITIQUE

DICTIONNAIRE DE L'OUTIL D'ÉVALUATION CRITIQUE D'UNE ÉTUDE ANALYTIQUE

Introduction

L'outil d'évaluation critique (OEC) d'une étude analytique a pour objet d'aider à l'évaluation de l'utilité des résultats d'une étude analytique unique. Le présent dictionnaire offre des données de base sur la façon de mener une évaluation critique d'une étude analytique et décrit les éléments de l'outil servant à attribuer un classement à la qualité de l'étude et aux données probantes qui en découlent. Il importe de comprendre que le dictionnaire n'offre pas d'explications en profondeur de tous les concepts et ne les illustre pas au moyen de tous les exemples possibles. Par conséquent, l'évaluateur doit faire preuve de jugement lorsqu'il interprète les critères et les applique à l'étude faisant l'objet de l'évaluation.

Les critères ici s'appliquent à toutes les études analytiques. Ces études peuvent porter sur une exposition particulière (p. ex. facteurs de risque, interventions, facteurs de protection ou démographiques) et/ou des résultats (p. ex. infections, maladies, comportements, effets ou état de santé) auprès de plus d'un groupe étudié. De plus amples précisions sont données lorsque les distinctions sont pertinentes à un plan d'étude particulier.

Utilisez cet outil pour évaluer les essais contrôlés randomisés et non randomisés, les études comparatives avant-après, les études en laboratoire, les études de cohorte, les études cas-témoins, les séries temporelles interrompues et les études non comparative avant-après. Un sommaire des caractéristiques de chaque plan d'étude figure au tableau 2, à la partie 2 de la présente trousse d'outils. Les études de modélisation mathématique ne sont pas abordées dans la présente trousse d'outils.

Évaluation critique de la validité d'une étude analytique

L'évaluation critique a pour principal objet d'évaluer la validité interne et la validité de la conclusion statistique. La *validité interne* désigne la mesure dans laquelle il est possible d'inférer (conclure) que l'exposition étudiée cause ou influence réellement le résultat étudié et que la relation entre les deux n'est pas artificielle ni la conséquence d'un autre facteur externe. La *validité de la conclusion statistique* signifie qu'il existe réellement entre l'exposition et le résultat une relation qui n'est pas totalement le fruit du hasard. Si la validité interne et la validité de la conclusion statistique sont fortes, l'association étudiée est solidement établie, et on peut considérer son application dans d'autres contextes. En l'absence de données probantes établissant une association, il devient inutile d'analyser l'application des conclusions ailleurs.

Consignes

Commencez l'évaluation critique en définissant le plan d'étude. Si vous ne pouvez pas définir le plan d'étude, discutez-en avec vos collègues et choisissez le plan d'étude qui s'en rapproche le plus. Classez chaque élément, à l'aide de l'outil, dans la catégorie fort, modéré ou faible, selon les critères décrits ici. Les critères ne sont pas toujours applicables au plan d'étude. À moins d'indication contraire, la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être respectés pour attribuer le classement en question. Quant aux critères d'un classement « faible », associé à l'énoncé « L'un ou l'autre des éléments suivants », l'élément devrait être classé dans la catégorie faible s'il ne possède qu'un seul critère.

Les critères à examiner sont inscrits en caractères gras italiques, et on donne une explication plus en profondeur. L'outil réserve un espace aux commentaires; l'évaluateur devrait ajouter des commentaires pour appuyer ses décisions et aider à définir les secteurs de préoccupations (p. ex. faiblesses ou limites majeures) et à déterminer s'ils contribueraient à une conclusion erronée au sujet de l'existence d'une association ou de sa force (surestimation ou sous-estimation d'un effet). Certaines études n'offrent pas assez de détails pour une évaluation adéquate de chaque critère. De telles études sont parfois non pas réellement faibles, mais simplement mal rédigées. Néanmoins, votre évaluation devrait porter uniquement sur ce qui est documenté dans le rapport d'étude et non pas sur vos hypothèses quant à l'étude. Remplissez le tableau de synthèse des données probantes durant le processus d'évaluation critique.

Si vous ne pouvez pas prendre une décision quant au classement d'un élément, écrivez un commentaire et discutez-en avec vos collègues. Les conclusions relatives à la qualité des données probantes sont généralement le fruit d'un consensus de groupe plutôt que d'une décision individuelle.

Filtrage des études à inclure avant l'évaluation critique

Avant de prendre la décision d'inclure ou non une étude, il faut la lire et brièvement déterminer ce qui a été fait. Si plus d'une question de recherche a été abordée ou que de multiples méthodes de recherche ont été employées, définissez les aspects qui se rapportent à votre **question clé** (voir le glossaire). Il faut noter qu'il est possible qu'un aspect d'une étude se rapporte à une question clé et qu'un autre aspect se rapporte à une différente question clé et que différentes méthodes de qualité variable soient employées. Une étude utilisée à l'appui de différentes conclusions doit être lue de nouveau en fonction de chaque question clé.

1. L'étude a-t-elle abordé une question bien ciblée qui se rapporte à la question clé?

Une question de recherche claire devrait être abordée. La population, l'intervention, le comparateur et les résultats étudiés devraient être précisés. Plus la question d'étude est ciblée, plus il est probable que les auteurs pourront y répondre.

La question de recherche de l'étude devrait se rapporter à la question clé des lignes directrices. Si l'étude ne se rapporte pas à la question clé, ses résultats n'aideront probablement pas à formuler des recommandations et il est inutile de consacrer du temps à sa lecture.

- Forte :** Question de recherche bien ciblée, lien solide avec la question clé.
- Modérée :** Question de recherche claire, assez bien ciblée et se rapportant à la question clé, quoiqu'elle ne permette pas toujours d'y répondre directement ou soit susceptible de donner lieu à des résultats exigeant une extrapolation pour aborder la question de recherche.
- Faible :** Question de recherche vague ou trop générale ou n'ayant aucun lien avec la question clé.

Décision de filtrage

Tirez une conclusion quant à la poursuite de l'évaluation critique ou au rejet de l'étude. Une étude classée dans la catégorie faible devrait être évaluée avec précaution ou ne pas être évaluée du tout.



Si on a l'intention de procéder à l'évaluation critique, il faut nommer le plan d'étude (à l'aide du bon algorithme à la partie 2) et remplir le tableau de synthèse des données probantes en plus de l'OEC dans le cadre du processus d'examen.

Évaluer la population à l'étude (échantillon) et la méthode d'échantillonnage

2. Les sujets sélectionnés pour participer à l'étude devraient représenter la population cible.

L'échantillon doit représenter la population cible pour que les inférences tirées des données soient valides pour ce groupe. On peut utiliser de multiples stratégies (p. ex. discussions personnelles, affiches et campagnes médiatiques) pour recruter des participants de diverses sources (p. ex. hôpitaux et collectivité). Le recrutement consiste à inviter et à encourager des participants potentiels à participer. Toutefois, la sélection consiste à choisir des participants parmi tous ceux qui sont disponibles. De plus, il faut déterminer comment les sujets ont été sélectionnés pour inclusion dans la base de données et si tous les groupes cibles seront inclus dans la base de données utilisée.

Dans certaines études, le recrutement ne s'applique pas; par exemple dans des études qui utilisent des bases de données existantes ou des études en laboratoire sans participants humains. Or, la sélection doit toujours être prise en considération. Les bases de données administratives ou autres appropriées susceptibles d'inclure la population étudiée devraient être utilisées dans le cadre d'études s'appuyant sur des bases de données. Dans le cas d'une étude en laboratoire avec participants humains, les méthodes de recrutement sont pertinentes.

| | |
|------------------|---|
| Forte : | De multiples stratégies ont été employées; participants recrutés ou sélectionnés à différents endroits ou dans différents groupes ou toute la population cible a été sélectionnée. Les participants (ou l'échantillon) possèdent clairement les caractéristiques ciblées ou la base de données administratives appropriée a été utilisée. |
| Modérée : | Les participants ont été sélectionnés d'une source unique susceptible d'avoir exclu des membres de la population cible. Les participants (ou l'échantillon) semblent posséder les caractéristiques ciblées. |
| Faible : | Les participants se sont proposés, sont des volontaires ou ce n'est pas clair à la lumière de la description, si les participants (ou l'échantillon) possèdent les caractéristiques ciblées ou ils ne possèdent clairement pas les caractéristiques. |

3. Justesse du contrôle du biais de sélection.

Il y a biais de sélection lorsqu'une erreur systématique se glisse dans le processus de sélection d'une population à étudier, surtout s'il existe des différences systématiques dans la relation entre l'exposition et le résultat du groupe témoin et du groupe d'intervention. Il y a biais de sélection si la relation entre l'exposition et le résultat diffère chez ceux qui participent à l'étude et ceux qui pourraient théoriquement le faire, mais ne le font pas. Cet écart importe seulement s'il est susceptible d'influencer l'association entre l'exposition et le résultat (p. ex. sujets plus susceptibles de profiter de l'intervention sont aussi plus susceptibles de participer).

Des méthodes et des critères de recrutement/sélection semblables devraient être appliqués au groupe d'intervention/exposé et au groupe témoin (ou aux cas et aux témoins dans une étude cas-témoin). Une proportion élevée ($\geq 80\%$) des personnes abordées doivent avoir accepté de participer et il ne doit pas y avoir de différences entre les groupes. Les caractéristiques de référence (autres que l'exposition ou le résultat étudié) devraient être semblables d'un groupe à l'autre (p. ex. âge, sexe, autres facteurs de risque connus et environnement), et il faut fournir des données à l'appui de la conclusion de similitudes.

L'échantillonnage aléatoire, c.-à-d. la sélection de participants à l'aide d'un processus aléatoire, atténue le biais de sélection si tout le monde accepte de participer. L'échantillonnage aléatoire ne s'applique pas aux essais contrôlés. La répartition aléatoire (randomisation) s'applique aux essais contrôlés et sera abordée dans une section ultérieure quant à la justesse du contrôle de la confusion. Dans le cas d'une étude en laboratoire, les méthodes de recrutement et les taux de participation ne s'appliquent pas, mais l'échantillon étudié (p. ex. matériel, microorganismes) devrait posséder des caractéristiques ciblées semblables.

| | |
|------------------|---|
| Forte : | Échantillonnage aléatoire utilisé et processus ou critères de recrutement ou de sélection semblables appliqués à tous les participants. Caractéristiques de référence semblables; $\geq 80\%$ ont accepté de participer; taux de participation semblables dans les deux groupes. |
| Modérée : | L'échantillonnage aléatoire n'a pas été utilisé, mais des processus ou des critères de recrutement ou de sélection semblables ont été appliqués à tous les participants. Les caractéristiques de référence étaient semblables; $\geq 80\%$ ont accepté de participer; taux de participation semblables dans les deux groupes. |
| Faible : | L'échantillonnage aléatoire n'a pas été utilisé. Les processus ou les critères de recrutement ou de sélection peuvent avoir différé dans le cas de certains participants. Certaines caractéristiques de référence n'étaient pas semblables; $< 80\%$ ont accepté de participer et/ou les taux de participation différaient d'un groupe à l'autre. |

Évaluer la validité interne

4. Justesse du contrôle du biais causé par une erreur de classification.

Il y a biais causé par une erreur de classification lorsque les participants ne sont pas associés aux bonnes catégories quant à leur exposition ou à leur résultat. ***Il faut appliquer de façon uniforme à tous les participants des définitions claires de l'exposition et du résultat, de même que les tests diagnostiques, et les mesures utilisées doivent être aussi objectives que possible. L'exposition doit clairement avoir eu lieu avant le résultat (lien temporel clair).*** Un lien temporel vague peut causer un problème dans le cadre des études rétrospective et cas-témoin. Dans le cas d'une STI, il devrait y avoir un moment clairement défini où est survenue l'intervention ou l'exposition.

Il devrait y avoir une quantité minimale de données manquantes ou inexactes, car les données manquantes ou les données erronées recueillies peuvent signifier qu'on ne connaît pas véritablement l'état d'exposition ou le résultat. Si les tests diagnostiques diffèrent, des participants du groupe témoin peuvent présenter le résultat ou avoir été exposés et l'ignorer. Les données manquantes peuvent causer un plus grand problème lorsqu'on travaille avec une base de données que lorsque les chercheurs recueillent leurs propres données. Le chercheur n'a aucun contrôle des données entrées dans une base de données, alors il ne peut pas en contrôler l'exactitude.

Une erreur de classification peut aussi survenir si l'intégrité de l'intervention était faible, ce qui risque d'arriver si : 1) tous les membres du groupe d'intervention n'ont pas subi la même intervention de la même façon (p. ex. en raison du non-respect de la méthode d'intervention ou d'un manque d'uniformité connexe); 2) les membres du groupe témoin peuvent avoir eu recours à l'intervention d'une autre source (p. ex. l'information a été obtenue d'un membre du groupe d'intervention ou d'une source externe ou il a été possible de recourir à l'autotraitement).

Une erreur de classification peut aussi survenir si on utilise des données agrégées comme mesure de résultats ou qu'il n'est pas clair si les sujets qui présentent le résultat ont été exposés (p. ex. intervention, facteur de risque et autre).

Forte : Intégrité de l'intervention forte, application de définitions claires, lien temporel clair, utilisation de mesures objectives pour établir l'état d'exposition ou le résultat et aucune donnée manquante ou inexacte.

Modérée : Intégrité de l'intervention forte, application de définitions claires, lien temporel clair, mais certaines données étaient manquantes ou des erreurs se sont glissées dans la mesure de l'exposition ou du résultat. Ces erreurs risquent d'avoir entraîné une erreur de classification touchant seulement quelques participants.

Faible : L'un ou l'autre des éléments suivants : Intégrité de l'intervention faible ou définitions vagues ou appliquées de façon disparate ou données manquantes ou erreurs s'étant glissées dans la mesure de l'exposition ou du résultat et risquant d'avoir causé une erreur de classification touchant un grand nombre de participants. Le lien temporel n'est pas toujours clair. Les résultats sont déclarés de façon agrégée, et on ignore si les résultats sont associés à l'exposition (p. ex. intervention, facteur de risque et autre).

5. Justesse du contrôle du biais d'information.

Un biais d'information peut être le résultat de procédures de collecte de données déficientes.

Les intervieweurs, par exemple, peuvent poser leurs questions différemment à différentes personnes ou ne pas interpréter l'information de la même façon. Les participants ressentant des effets indésirables sont susceptibles de ne pas se souvenir de leurs expériences de la même façon que les sujets qui n'ont pas ces effets (biais de rappel) ou les participants peuvent donner des réponses correctes sur le plan social ou politique ou dire ce qu'ils croient que le chercheur veut entendre (désirabilité sociale ou biais de déclaration). Parmi les stratégies d'atténuation de ces biais, on compte le fait de tenir l'évaluateur dans l'ignorance du statut d'intervention ou d'exposition des participants, des protocoles normalisés de collecte de données, la formation des évaluateurs pour promouvoir la fidélité interévaluateurs et le respect des protocoles, la formulation des questions et la prise de mesures (p. ex. anonymat et création d'un rapport) pour mettre les participants plus à l'aise de répondre honnêtement à des questions difficiles. Le biais de rappel est problématique dans le cadre d'études cas-témoins et d'études de cohorte rétrospectives.

Maintenir l'insu est une méthode utile surtout lorsqu'on sait que le groupe auquel le participant appartient peut avoir une incidence sur le résultat mesuré (p. ex. détresse psychologique) ou le respect du protocole (p. ex. perte de poids). Maintenir l'insu n'est pas nécessairement pertinent dans certaines études en laboratoire.

Il y a une différence entre la collecte de données particulières à l'étude (p. ex. entrevues) et la collecte de données cliniques de routine (p. ex. lorsque les données sont recueillies pour l'étude à la lumière d'un examen des dossiers ou des renseignements obtenus à des fins cliniques). Il est généralement raisonnable de croire que les professionnels des soins de santé, comme les médecins et les infirmières, ont suivi une formation appropriée pour la collecte de données cliniques de routine, comme des antécédents, des examens physiques ou des échantillons prélevés en laboratoire clinique, et que le personnel de laboratoire a suivi une formation appropriée pour traiter les échantillons en laboratoire. L'évaluation critique vise principalement la méthode de collecte de données pour l'étude.

- Forte :** Les évaluateurs ont été tenus dans l'ignorance quant au groupe des participants, ont suivi une formation sur les procédures de collecte de données et les ont clairement respectées. Des stratégies pour minimiser les biais associés aux procédures et mesures de collecte de données ou la formulation des questions ont été utilisées. Les patients étaient tenus dans l'ignorance ou non, mais n'ont eu aucune incidence sur les données recueillies.
- Modérée :** Les évaluateurs n'ont pas été tenus dans l'ignorance quant au groupe des participants, mais ils ont suivi une formation sur les procédures de collecte de données et les ont probablement respectées. Des stratégies pour réduire les biais associés aux procédures et mesures de collecte de données ou la formulation des questions ont été utilisées. Les patients n'ont pas été tenus dans l'ignorance, et ils auraient pu avoir une incidence sur les données recueillies.
- Faible :** Les évaluateurs n'ont pas été tenus dans l'ignorance quant au groupe des participants, et on ignore s'ils ont suivi une formation sur les procédures de collecte de données ou les ont suivies. On ignore si les stratégies suffisaient à réduire le biais des réponses lié aux procédures et mesures de collecte de données ou à la formulation des questions. Les patients n'ont pas été tenus dans l'ignorance, ce qui a clairement eu une incidence sur les données recueillies.

6. Validité et fiabilité des instruments de collecte de données

Les instruments employés pour recueillir les données devraient être valides et fiables.

On devrait distinguer les instruments (p. ex. guide d'entrevue, questionnaire, formulaire d'extraction des données, méthodes en laboratoire, etc.) des méthodes de collecte de données (p. ex. entrevue, sondage, examen des dossiers). La validité tient au fait que l'instrument mesure ce qu'il est censé mesurer, et la fiabilité tient à l'uniformité du procédé. On peut employer des méthodes d'essai particulières pour évaluer la validité et la fiabilité des instruments, et on accorde une plus grande confiance à ceux ayant fait l'objet de ces essais.

- Forte :** Les outils sont réputés valides et fiables ou on l'a démontré.
- Modérée :** Il n'y a eu aucune tentative visant à évaluer la validité et la fiabilité des outils, mais on peut croire à la validité du contenu du fait de la nature des questions posées et de la participation d'experts à l'élaboration des outils.
- Faible :** Il n'y a eu aucune tentative en vue d'évaluer la validité et la fiabilité, et rien ne donne à penser que ces caractéristiques sont présentes.

7. Justesse de la conservation et du suivi.

Il importe que les participants terminent l'étude pour que des données adéquates soient disponibles relativement à tous les résultats étudiés chez les participants des deux groupes. Une étude incomplète peut être le résultat d'un participant qui a ressenti des effets indésirables, voire qui est mort, ou qui se portait bien et n'est pas revenu pour être évalué. On considère qu'un *suivi est impossible* si on ne peut joindre le participant, et la raison de l'étude incomplète est donc inconnue. Par conséquent, **on devrait tout tenter pour découvrir pourquoi un participant n'a pas terminé l'étude**. Le taux d'abandon peut compromettre les conclusions finales quant à l'association entre l'exposition et le résultat si les participants n'ont pas terminé l'étude parce qu'ils sont décédés, avaient trop d'effets secondaires ou surtout si les raisons se rattachaient à l'une des variables étudiées. **Idéalement, une grande proportion de participants devraient terminer l'étude, il ne devrait pas y avoir de différence dans le taux d'abandon des groupes et les raisons du défaut de terminer l'étude ne devraient pas se rattacher à l'exposition étudiée**. Le suivi impossible peut être un problème important dans le cadre de toute étude prospective, dont les études de cohorte et les essais contrôlés ou les études comparatives avant-après dont la période de suivi est longue. En général, ce n'est pas un problème dans le cadre d'une étude de STI, qui prévoit des périodes d'évaluation adéquates, mais il faut l'évaluer.

Les taux d'abandon devraient être interprétés en fonction du résultat étudié. Dans certaines études, il peut y avoir une distinction entre le parachèvement de l'étude et le parachèvement d'une thérapie auprès des patients qui ne respectent pas le protocole. De tels patients peuvent continuer à être suivis aux fins de l'étude. Dans le cadre d'une étude en laboratoire, un échantillon compromis, mal manipulé ou non viable peut être considéré comme impossible à suivre.

| | |
|------------------|--|
| Forte : | Une proportion élevée (> 90 %) des participants ont terminé l'étude, il n'y a pas eu de différence dans le taux d'abandon entre les groupes et les raisons de l'abandon n'étaient pas liées à l'exposition. |
| Modérée : | Une proportion assez élevée (\geq 80 %) des participants a terminé l'étude, il y avait peu de différence dans le taux d'abandon entre les groupes et les raisons de l'abandon n'étaient pas liées à l'exposition. |
| Faible : | L'un ou l'autre des éléments suivants : Une faible proportion (moins de 80 %) des participants ont terminé l'étude, il y avait de grandes différences dans le taux d'abandon entre les groupes ou les raisons de l'abandon étaient liées à l'exposition. |

Évaluer le contrôle de la confusion

Les variables confusionnelles sont des variables susceptibles de déformer le lien entre l'exposition et le résultat ou de présenter une explication plausible du lien observé (c.-à-d. les résultats peuvent être également causés par la variable confusionnelle et non pas par l'exposition étudiée). Ainsi, il est crucial de contrôler la confusion pour pouvoir conclure que l'association observée est réellement la conséquence de l'exposition étudiée. Parmi les variables confusionnelles courantes, on peut mentionner l'âge, le sexe et le contexte, mais, en réalité, les variables confusionnelles varient selon l'association étudiée; elles devraient être mesurées comme valeurs de base dans les deux groupes pour évaluer les similitudes.

8. Comparabilité du groupe témoin et du groupe d'intervention/exposé.

Un groupe témoin ou de comparaison permet l'évaluation et le contrôle des variables confusionnelles potentielles. Les participants du groupe témoin devraient être comparables à ceux du groupe d'intervention (p. ex. dans un ECR) ou à ceux du groupe avec exposition (p. ex. dans une étude de cohorte) à l'exception de l'intervention ou de l'exposition respectivement. Tous les groupes d'une STI devraient être semblables, sauf en ce qui concerne l'exposition. Il est difficile d'assurer la similitude des groupes lorsque le groupe témoin n'est pas évalué simultanément. De façon analogue, dans le cadre d'une étude cas-témoin, les cas et les témoins devraient présenter des caractéristiques semblables à part l'exposition et le résultat étudié. Trouver des participants convenables pour un groupe témoin est très difficile dans des études cas-témoin. Dans le cadre d'une expérience en laboratoire, la comparaison est effectuée avec une condition témoin plutôt qu'un groupe témoin.

Sans égard au plan d'étude, la collecte de données devrait avoir lieu simultanément pour éliminer la possibilité d'autres changements pouvant expliquer les résultats (p. ex. changements sur le plan de l'environnement ou des pratiques au fil du temps). L'utilisation des sujets comme mesure de contrôle est une méthode d'atténuation de la confusion, mais cela ne permet pas d'évaluer de multiples facteurs de risque ou des variables confusionnelles autres que les caractéristiques individuelles.

- Forte :** Les deux groupes étaient semblables à la base sur le plan des caractéristiques clés susceptibles d'avoir une incidence sur le résultat, et le groupe témoin a été évalué en même temps que le groupe d'intervention. Dans le cadre d'une étude cas-témoin, les témoins sont appropriés pour les cas.
- Modérée :** Les deux groupes étaient comparables, à l'exception de différences mineures peu susceptibles de compromettre le résultat. Dans le cadre d'une étude cas-témoin, les témoins sont appropriés pour les cas.
- Faible :** L'un ou l'autre des éléments suivants : Il n'y avait pas de groupe témoin (même si les participants font office de témoin) ou les groupes ne peuvent pas être considérés comme étant comparables (il y avait de grandes différences entre eux) ou la similitude des groupes n'a pas été évaluée.

9. Justesse du contrôle des grandes variables confusionnelles

La répartition aléatoire dans des groupes (randomisation) répartit les variables confusionnelles inconnues de façon égale dans les groupes et est une importante stratégie de contrôle de la confusion. **Le processus de répartition aléatoire donne à chaque participant à l'étude une possibilité égale d'être réparti dans un groupe ou un autre.** Voici des exemples de processus de répartition aléatoire : utilisation de générateurs de chiffres aléatoires et tirer à pile ou face. Il faut noter que la répartition aléatoire à des groupes permet de contrôler la confusion, tandis que l'échantillonnage aléatoire (sélection aléatoire de participants) n'a pas d'incidence sur la confusion, mais contribue à la possibilité de généraliser les résultats.

Lorsque la répartition aléatoire (randomisation) aux groupes n'est pas effectuée, d'autres stratégies peuvent être employées. L'appariement de cas et de témoins en fonction des variables confusionnelles connues (correspondance individuelle ou par groupe) et l'utilisation d'une analyse statistique appropriée (p. ex. modélisation ou analyse stratifiée) permettent aussi de contrôler la confusion. Dans le cas d'une STI, il faut mener au moins trois évaluations avant l'exposition (p. ex. intervention, facteur de risque ou autre) et trois évaluations après pour bien évaluer les tendances au fil du temps (source de confusion). Les auteurs devraient aussi aborder d'autres facteurs susceptibles d'influencer le résultat, p. ex. saison, tendances séculaires ou autres interventions ou activités sans lien avec l'étude. Il est impossible de définir toutes les variables confusionnelles, mais **les chercheurs devraient déterminer les principales variables confusionnelles potentielles, évaluer leur présence et les contrôler, sans égard au type de plan**. Dans le cadre d'une expérience en laboratoire, la répartition aléatoire ne s'applique pas, qu'il y ait des participants humains ou non. Les chercheurs assurent le contrôle des variables confusionnelles potentielles et s'y adaptent.

- | |
|---|
| <p>Forte : Il y a eu répartition aléatoire à l'aide d'un processus approprié. Sinon, un appariement approprié ou l'analyse statistique ou les conditions en laboratoire ont permis le contrôle adéquat de la confusion. Les grandes variables confusionnelles ont été examinées.</p> <p>Modérée : Il y a eu répartition systématique aux groupes, mais pas de véritable répartition aléatoire, ou le processus de répartition aléatoire était vague ou inadéquat ou il n'y avait pas d'appariement approprié, mais une analyse statistique a permis un contrôle adéquat de la confusion ou les conditions expérimentales en laboratoire ont permis le contrôle partiel de la confusion et les grandes variables confusionnelles ont été examinées.</p> <p>Faible : Il n'y a pas eu de répartition aléatoire pour distribuer les sujets dans les groupes, ni d'appariement approprié et l'analyse statistique ou les conditions expérimentales en laboratoire n'ont pas donné lieu au contrôle de la confusion et/ou les grandes variables confusionnelles n'ont pas été examinées.</p> |
|---|

Éthique

10. Justesse de la conduite éthique

Sans égard au plan, il faut avoir pris les mesures appropriées pour protéger les participants, surtout les groupes vulnérables, du danger, de l'exploitation et de la coercition, ainsi que leur droit à l'autodétermination, à la divulgation complète, au traitement équitable et à la protection des renseignements personnels. Le consentement éclairé est l'une des grandes stratégies visant à protéger les droits. Le chercheur qui déclare que son étude a été approuvée par un conseil d'examen éthique institutionnel sans donner plus de détails doit avoir fait examiner et approuver cette déclaration au préalable. Il faut noter que, selon l'Énoncé de politique des trois Conseils, l'évaluation de programmes et les études de surveillance n'exigent pas une approbation éthique, mais il faut tout de même prendre des mesures appropriées pour protéger les participants et leurs droits. Les investigations de santé publique, tel que les traçabilités et les investigations d'exposition ne nécessitent pas de revue éthique. Si une étude n'est pas menée de façon éthique, les données peuvent être biaisées. On considère aussi qu'il n'est pas éthique d'utiliser les résultats de ces études. Voici des exemples de détails adéquats ou suffisants liés à la conduite éthique à l'égard de l'utilisation de données existantes : élimination des identificateurs des données, obtention de la permission du responsable des données et utilisation d'une base de données gouvernementale ou institutionnelle.

L'éthique de la recherche s'attache aussi à l'influence indue de promoteurs ou d'autres intervenants sur les méthodes ou la déclaration des résultats de façon à obtenir une conclusion favorable.

- | | |
|--|---|
| <p>Forte :</p> <p>Modérée :</p> <p>Faible :</p> | <p>La recherche a été approuvée par un conseil d'examen éthique approprié ou on a fourni suffisamment de détails pour garantir la conduite éthique. Le rapport de recherche n'a pas été influencé par un organisme de financement, un promoteur ou un conflit d'intérêts.</p> <p>Ne s'applique pas.</p> <p>Une quantité insuffisante de détails est fournie et il est impossible de tirer une conclusion quant à la conduite éthique. La possibilité que le rapport de recherche ait été influencé par un organisme de financement, un promoteur ou un conflit d'intérêts n'a pas pu être exclue.</p> |
|--|---|

Évaluer l'analyse

11. Justesse et interprétation des tests statistiques

Les tests statistiques employés dans l'analyse devraient être adaptés au type et à la catégorie de données et correctement appliqués. Par exemple, la régression convient au calcul du rapport de cotes (RC) lorsque le contrôle de variables confusionnelles multiples est nécessaire, tandis que le test t convient à la comparaison de la moyenne entre les deux groupes. Il ne suffit pas de mener une analyse univariée lorsque les données suffisent à mener une analyse multivariée si la taille de l'échantillon est suffisante. Voir la « synthèse des tests statistiques communs » à l'annexe A, tableau 5 de la présente trousse d'outils.

Les critères (p. ex. $\alpha=0,05$) qui établissent la signification statistique devraient être clairs et convenables. Il faut donner la valeur p et bien l'interpréter (p. ex. le résultat était statistiquement significatif si la valeur p était inférieure à α , p. ex. $p<0,05$). **Il faut aussi bien interpréter les intervalles de confiance (IC) lorsqu'ils sont fournis** (p. ex. un IC d'un RC qui inclut la valeur 1 indique qu'il n'y a pas de différence entre les deux groupes).

Forte : Les tests statistiques convenaient à la catégorie de données et aux hypothèses éprouvées. Les valeurs de probabilité et les intervalles de confiance ont bien été interprétés.

Modérée : De simples tests ont bien été utilisés, mais les données exigeaient des tests plus pointus, et le contrôle de la confusion était limité.

Faible : Les tests étaient inadaptés aux données ou on n'a pas fourni de renseignements quant aux tests utilisés. Les résultats n'ont pas bien été interprétés.

12. Puissance et taille de l'échantillon

La puissance renvoie à la capacité d'une étude de détecter une différence statistiquement significative entre des groupes lorsqu'une telle différence existe. Une puissance de 80 % est habituellement considérée comme étant suffisante. Plus l'échantillon est grand ou plus la différence entre les deux groupes est grande, plus il est facile de détecter une différence statistiquement significative. Il importe de noter que, si une différence significative a été établie, l'étude est suffisamment puissante, peu importe la taille de l'échantillon. **Toutefois, les études dont la taille de l'échantillon est insuffisante et, par conséquent, dont la puissance est insuffisante ne permettent pas de déterminer si les résultats sont purement le fruit du hasard.**

Forte : Des différences significatives ont été établies; par conséquent, la taille de l'échantillon était suffisante ou aucune différence significative n'a été constatée, mais le chercheur a déclaré que la puissance permettait d'établir une telle différence.

Modérée : Aucune différence significative n'a été établie, et le chercheur a déclaré que la puissance de l'étude était insuffisante. La taille de l'échantillon semblait raisonnable selon le plan et les questions de recherche; p. ex. justification par d'autres études.

Faible : Aucune différence significative n'a été établie, l'échantillon était petit et les chercheurs n'ont pas abordé la justesse de la puissance de l'étude.

Évaluer l'applicabilité

L'évaluation de l'applicabilité (possibilité de généralisation et faisabilité) n'est pas nécessairement pertinente dans le cadre de toutes les études, surtout celles qui portent sur un facteur de risque. L'évaluation de l'applicabilité n'a aucune incidence sur les décisions liées à la qualité de l'étude. L'évaluateur peut décider d'analyser l'applicabilité en fonction de l'objet de son évaluation de l'étude. L'interprétation des critères d'applicabilité devrait se faire en consultation avec des collègues.

13. Les résultats peuvent-ils être généralisés dans la population locale?

Il n'est pas nécessairement possible de généraliser les données recueillies auprès d'un groupe à d'autres groupes s'il existe de grandes différences d'un groupe à l'autre.

L'échantillonnage aléatoire, la sélection d'échantillons d'un groupe diversifié et le fait de s'assurer que l'échantillon représente adéquatement d'autres groupes étudiés contribuent tous à accroître la possibilité de généralisation.

Forte : Les caractéristiques de la population à l'étude étaient très semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

Modérée : Les caractéristiques de la population à l'étude étaient quelque peu semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

Faible : Les caractéristiques de la population à l'étude n'étaient pas du tout semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

14. Faisabilité de la mise en œuvre

La faisabilité de la mise en œuvre d'une intervention varie selon le contexte et peut dépendre de l'accessibilité aux ressources (p. ex. fonds, personnel adéquat, volonté politique et environnement physique). On peut aussi prendre en considération le caractère acceptable aux yeux des patients, du personnel et d'autres intervenants.

Forte : L'intervention à l'étude est fortement susceptible d'être facilement mise en œuvre dans d'autres contextes.

Modérée : L'intervention est plutôt susceptible d'être facilement mise en œuvre dans d'autres contextes ou aucune intervention n'a été étudiée, mais l'exposition étudiée est très susceptible de déclencher une intervention qui pourra facilement être mise en œuvre.

Faible : Il est peu probable que l'intervention soit facilement mise en œuvre dans d'autres contextes ou aucune intervention n'a été étudiée, et l'exposition étudiée n'est pas très susceptible de donner lieu à une intervention qui pourra facilement être mise en œuvre.

Décision relative à la qualité de l'étude

15. Résumer les résultats de l'évaluation critique

DÉTERMINER LA FORCE DU PLAN D'ÉTUDE

Décision relative à la qualité de l'étude

Examinez les classements attribués aux éléments 2 à 12 à évaluer.

Classez la qualité dans la catégorie ÉLEVÉE si vous avez attribué un classement fort à la plupart des éléments à évaluer, voire tous, et qu'aucun élément n'était faible. En plus, il n'y a pas d'importants obstacles à la validité interne de l'étude ni à la capacité de tirer une conclusion quant à l'association claire entre l'exposition et le résultat étudié.

Classez la qualité dans la catégorie MOYENNE si les éléments 4 et/ou 11 ont été classés au moins dans la catégorie modérée et que les autres éléments à évaluer classés comme étant faibles ou modérés ne suffisent pas à compromettre la validité interne de l'étude. Aussi, ces autres éléments ne doivent pas compromettre la capacité de tirer une conclusion selon laquelle il y a probablement une association entre l'exposition et le résultat étudié.

Classez la qualité dans la catégorie FAIBLE si les éléments 4 et/ou 11 ont été classés dans la catégorie faible ou si d'autres éléments classés dans cette catégorie suffisent à compromettre la capacité d'exclure d'autres explications quant aux constatations et de tirer une conclusion relative à l'association entre l'exposition et le résultat étudié.

Décision relative au caractère direct des données probantes présentées dans l'étude

Tirez une conclusion relative au caractère direct des données probantes :

- Les **données probantes directes** proviennent des études portant précisément sur l'association étudiée.
- L'**extrapolation** est l'inférence tirée d'études qui portaient sur une question différente mais connexe ou la même question, mais dans des conditions artificielles (p. ex. certaines études en laboratoire).

Indiquez, dans la mesure du possible, toute recommandation provisoire à des fins pratiques tirée de l'étude, en n'oubliant pas qu'une recommandation doit être fondée sur le corpus de données probantes et non pas sur une seule étude et que, globalement, les avantages doivent l'emporter sur tout préjudice ou coût. Consultez la section Formuler une recommandation et examinez tout le corpus de données probantes (pas seulement une étude) et discutez-en avec vos collègues avant de formuler une recommandation.



Remplir le tableau des données probantes pour s'assurer qu'il contient les principaux détails liés à l'intervention ou à l'exposition, à l'échantillon, aux méthodes et aux résultats. Inclure la « force du plan », le « caractère direct des données probantes » et vos conclusions au sujet de la qualité de l'étude et du caractère direct des données probantes.

OUTIL D'ÉVALUATION CRITIQUE D'UNE ÉTUDE ANALYTIQUE

Question clé : _____

Auteur : _____ Année : _____ Numéro de référence : _____

Titre : _____

Évaluateur : _____ Date : _____

Voir le dictionnaire de l'outil d'évaluation critique d'une étude analytique pour obtenir la liste complète des critères.

Les critères ne s'appliquent pas tous aux études. À moins d'indication contraire (par l'énoncé « l'un ou l'autre des éléments suivants »), la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être présents pour que l'on attribue à l'élément le classement en question.

| Sélectionner le plan d'étude | | | | | | | | | |
|------------------------------|------|-------------|-------|---------|---------|------------|--------------------|--------|----------------------|
| Forte | | | | Modérée | | | | Faible | |
| ECR | ECNR | Laboratoire | ECAA* | ECAA* | Cohorte | Cas-témoin | STI* (adéquate) | ENCAA | STI* (inadéquate) |
| | | | | | | | | | |

* Voir le Tableau 1 et la légende de « l'algorithme pour nommer le type d'étude analytique » pour une décision reliée à l'ECAA ou à la STI.

| Question de filtrage | | | |
|---------------------------------|--|---|--|
| | Forte | Modérée | Faible |
| 1. Question de recherche | Bien ciblée. Lien solide avec la question clé. <input type="checkbox"/> | Assez bien ciblée. Lien avec la question clé. <input type="checkbox"/> | Vague ou trop générale. Aucun lien avec la question clé. <input type="checkbox"/> |
| Commentaires : | | | |

| Décision relative au filtrage | | |
|--|----|-------------------------------------|
| <input type="checkbox"/> Rejeter (si faible) | OU | <input type="checkbox"/> Poursuivre |

| Évaluer la population de l'étude (échantillon) et la méthode d'échantillonnage | | | |
|--|--|--|---|
| | Forte | Modérée | Faible |
| 2. Participants de l'étude représentatifs de la population cible | Utilisation de stratégies de recrutement multiples. Recrutement ou sélection d'une variété d'endroits ou inclusion de tous les membres de la population cible. Les participants (ou l'échantillon en laboratoire) possèdent les caractéristiques ciblées ou on a utilisé une base de données appropriée. <input type="checkbox"/> | Recrutement ou sélection de participants d'une source unique susceptible d'avoir exclu des membres de la population cible. Les participants (ou l'échantillon) semblent posséder les caractéristiques ciblées. <input type="checkbox"/> | Les participants se sont proposés ou sont volontaires. Les participants (ou l'échantillon) ne possèdent pas les caractéristiques ciblées ou ce n'est pas clair si c'est le cas. <input type="checkbox"/> |
| 3. Justesse du contrôle du biais de sélection | Utilisation de l'échantillonnage aléatoire. Processus de recrutement ou de sélection semblable appliqué partout, Caractéristiques de référence semblables, taux de participation $\geq 80\%$ dans chaque groupe. <input type="checkbox"/> | Pas d'échantillonnage aléatoire. Processus de recrutement ou de sélection semblable utilisé partout, Caractéristiques de référence semblables, taux de participation $\geq 80\%$ dans chaque groupe. <input type="checkbox"/> | Pas d'échantillonnage aléatoire. Le processus de recrutement ou de sélection et certaines caractéristiques de référence peuvent avoir varié. Taux de participation inférieur à 80% ou différent d'un groupe à l'autre. <input type="checkbox"/> |
| Commentaires : | | | |

| Évaluer la validité interne | | | |
|--|---|---|---|
| | Forte | Modérée | Faible |
| 4. Justesse du contrôle du biais causé par une erreur de classification | Intégrité de l'intervention forte et définitions claires appliquées. Lien temporel clair. Utilisation de mesures objectives pour établir l'état d'exposition ou le résultat. Aucune donnée manquante ou inexacte. <input type="checkbox"/> | Intégrité de l'intervention forte et définitions claires. Lien temporel clair. Certaines données manquantes ou des erreurs se sont glissées dans la mesure de l'exposition ou du résultat susceptibles d'avoir entraîné une erreur de classification chez seulement quelques participants. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Faible intégrité de l'intervention et définitions vagues. Lien temporel vague. Certains résultats déclarés de façon agrégée et impossible de déterminer avec certitude si les sujets ont subi l'intervention. Données manquantes ou des erreurs se sont glissées dans la mesure de l'exposition ou du résultat susceptibles d'avoir entraîné des erreurs de classification chez beaucoup de participants. <input type="checkbox"/> |
| 5. Justesse du contrôle du biais d'information | Évaluateurs tenus dans l'ignorance, ayant suivi une formation sur la collecte de données et ont clairement respectés les procédures. Biais associés aux procédures et mesures de collecte de données minimales. Les patients ont été tenus dans l'ignorance ou non, mais cela n'avait aucune incidence sur les données recueillies. <input type="checkbox"/> | Les évaluateurs n'ont pas été tenus dans l'ignorance, mais ils ont suivi une formation sur la collecte de données et ont probablement respectés les procédures. Biais associés aux procédures et mesures de collecte de données réduits. Les participants n'ont pas été tenus dans l'ignorance, mais ils auraient pu avoir une incidence sur les données recueillies. <input type="checkbox"/> | Les évaluateurs n'ont pas été tenus dans l'ignorance et on ignore s'ils ont suivi une formation sur la collecte de données ou respecté des méthodes particulières. On ne sait pas si le biais a suffisamment été réduit. Les patients n'ont pas été tenus dans l'ignorance, et ce qui a clairement eu une incidence sur les données. <input type="checkbox"/> |

| | | | |
|--|--|---|--|
| 6. Validité et fiabilité des instruments de collecte de données | Outils réputés valides et fiables ou cela a été démontré. <input type="checkbox"/> | Aucune tentative d'évaluer la validité et la fiabilité des outils. On peut croire à la validité du contenu en fonction des questions posées et de la participation d'experts. <input type="checkbox"/> | Aucune tentative d'évaluer la validité et la fiabilité des outils. Rien ne donne à penser que c'est le cas. <input type="checkbox"/> |
| 7. Justesse de la conservation et du suivi | > 90 % des participants ont terminé l'étude. Taux d'abandon semblable entre les groupes et raisons n'ayant aucun lien avec l'exposition. <input type="checkbox"/> | ≥ 80 % des participants ont terminé l'étude. Peu de différence dans le taux d'abandon entre les groupes et raisons n'ayant aucun lien avec l'exposition. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : < 80 % des participants ont terminé l'étude. Différences majeures dans le taux d'abandon entre les groupes ou raisons de l'abandon étaient liées à l'exposition. <input type="checkbox"/> |

| Évaluer le contrôle de la confusion | | | |
|--|--|---|--|
| | Forte | Modérée | Faible |
| 8. Comparabilité du groupe témoin et du groupe d'intervention | Les groupes étaient semblables à la base et ont fait l'objet d'une évaluation simultanée. On a employé des témoins appropriés dans le cadre d'une étude cas-témoin. <input type="checkbox"/> | Les groupes étaient comparables à la base avec des différences minimales. Témoins appropriés dans le cadre d'une étude cas-témoin. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Aucun groupe témoin ou de grandes différences entre les groupes ou non-évaluation de la similitude des groupes. <input type="checkbox"/> |
| 9. Justesse du contrôle des grandes variables confusionnelles | Répartition aléatoire appropriée pour distribuer les participants dans les groupes ou appariement approprié, analyse statistique ou conditions en laboratoire adéquates pour le contrôle de la confusion. Grandes variables confusionnelles examinées. <input type="checkbox"/> | Répartition aléatoire vague ou inadéquate ou appariement inapproprié, mais analyse statistique ayant permis le contrôle de la confusion ou conditions en laboratoire ayant permis le contrôle partiel de la confusion. Grandes variables confusionnelles examinées. <input type="checkbox"/> | Aucune répartition aléatoire pour distribuer les participants dans les groupes ni d'appariement approprié. L'analyse statistique ou les conditions en laboratoire n'ont pas donné lieu au contrôle de la confusion. Les grandes variables confusionnelles n'ont pas été examinées. <input type="checkbox"/> |
| Commentaires : | | | |

| Éthique | | | |
|---|--|--------------------|---|
| | Forte | Modérée | Faible |
| 10. Justesse de la conduite éthique <input type="checkbox"/> Ne s'applique pas (voir dictionnaire) | L'étude a été approuvée par un conseil d'examen éthique approprié ou il y avait assez de détails établissant que la conduite était éthique. Le rapport de recherche n'a pas été influencé. <input type="checkbox"/> | Ne s'applique pas. | La quantité de détails fournis n'était pas suffisante pour tirer une conclusion relative à la conduite éthique. La probabilité que le rapport de recherche a été influencé ne peut être exclue. <input type="checkbox"/> |
| Commentaires : Remarque : Selon la politique des trois Conseils, l'évaluation de programmes et les études de surveillance ne doivent pas faire l'objet d'une approbation éthique. | | | |

| Évaluer l'analyse | | | |
|--|--|---|---|
| | Forte | Modérée | Faible |
| 11. Justesse et interprétation des tests statistiques (Voir tableau 5) | Les tests statistiques conviennent à la catégorie de données et aux hypothèses éprouvées. Les valeurs de probabilité et les intervalles de confiance ont bien été interprétés. <input type="checkbox"/> | Des tests simples ont bien été utilisés, mais les données exigeaient des tests plus pointus. Le contrôle de la confusion était limité. <input type="checkbox"/> | Les tests ne convenaient pas aux données ou l'on n'a pas donné d'information sur les tests utilisés. Résultats mal interprétés. <input type="checkbox"/> |
| 12. Puissance et taille de l'échantillon | On a relevé des différences significatives; par conséquent, la taille de l'échantillon était suffisante ou on n'a pas relevé de différences significatives, mais les chercheurs ont déclaré que la puissance était suffisante. <input type="checkbox"/> | Aucune différence significative n'a été relevée, et les chercheurs ont déclaré que la puissance de l'étude n'était pas suffisante. La taille de l'échantillon semblait raisonnable. <input type="checkbox"/> | On n'a pas relevé de différences significatives, et la taille de l'échantillon était restreinte. La justesse de la puissance de l'étude n'a pas été mentionnée. <input type="checkbox"/> |
| Commentaires : | | | |

| Évaluer l'applicabilité | | | |
|--|---|---|---|
| | Forte | Modérée | Faible |
| | <input type="checkbox"/> Ne s'applique pas | | <input type="checkbox"/> N'a pas été évaluée |
| 13. Possibilité de généralisation des résultats | Les caractéristiques de la population à l'étude étaient très semblables à celles du groupe auprès duquel on souhaite généraliser les résultats. <input type="checkbox"/> | Les caractéristiques de la population à l'étude étaient quelque peu semblables à celles du groupe auprès duquel on souhaite généraliser les résultats. <input type="checkbox"/> | Les caractéristiques de la population à l'étude n'étaient pas du tout semblables à celles du groupe auprès duquel on souhaite généraliser les résultats. <input type="checkbox"/> |
| 14. Faisabilité de la mise en œuvre | L'intervention est fortement susceptible d'être facilement mise en œuvre dans d'autres contextes. <input type="checkbox"/> | L'intervention est plutôt susceptible d'être facilement mise en œuvre ou l'exposition est très susceptible d'entraîner une intervention qui pourra facilement être mise en œuvre. <input type="checkbox"/> | L'intervention est peu susceptible d'être facilement mise en œuvre ou l'exposition est peu susceptible d'entraîner une intervention qui pourra facilement être mise en œuvre. <input type="checkbox"/> |
| Commentaires : | | | |

Conclusion globale

15. Synthétisez les résultats de l'évaluation critique et remplissez le tableau de synthèse des données probantes. Veuillez noter que vous ne pouvez pas formuler une recommandation en fonction d'une seule étude.

a) Déterminer la force du plan d'étude

(voir « Sélectionner le plan d'étude » au début de cet outil)

Forte Modérée Faible

b) Décision relative à la qualité de l'étude

Examinez les classements que vous avez attribués aux éléments 2 à 12 à évaluer et définissez le classement approprié pour la qualité.

Élevée Moyenne Faible

Classez la qualité dans la catégorie ÉLEVÉE si vous avez attribué un classement fort à la plupart des éléments à évaluer, voire tous, et qu'aucun élément n'était faible. En plus, il n'y a pas d'importants obstacles à la validité interne de l'étude ni à la capacité de tirer une conclusion quant à l'association claire entre l'exposition et le résultat étudié.

Classez la qualité dans la catégorie MOYENNE si les éléments 4 et/ou 11 ont été classés au moins dans la catégorie modérée et que les autres éléments à évaluer classés comme étant faibles ou modérés ne suffisent pas à compromettre la validité interne de l'étude. Aussi, ces autres éléments ne doivent pas compromettre la capacité de tirer une conclusion selon laquelle il y a probablement une association entre l'exposition et le résultat étudié.

Classez la qualité dans la catégorie FAIBLE si les éléments 4 et/ou 11 ont été classés dans la catégorie faible ou si d'autres éléments classés dans cette catégorie suffisent à compromettre la capacité d'exclure d'autres explications quant aux constatations et de tirer une conclusion relative à l'association entre l'exposition et le résultat étudié.

c) Décision relative au caractère direct des données probantes

Direct Extrapolation

Commentaires :

DICTIONNAIRE DE L'OUTIL D'ÉVALUATION CRITIQUE D'UNE ÉTUDE DESCRIPTIVE

Introduction

L'objet de l'outil d'évaluation critique (OEC) d'une étude descriptive est d'aider à évaluer l'utilité des résultats d'une seule étude descriptive. Le présent dictionnaire a pour objet de donner des renseignements de base au sujet des limites de ces études et de décrire les éléments de l'outil pour aider les évaluateurs à attribuer un classement à la qualité de l'étude. Il importe de comprendre que le dictionnaire n'offre pas une explication en profondeur de tous les concepts et ne les illustre pas au moyen de tous les exemples possibles. Par conséquent, l'évaluateur doit faire preuve de jugement lorsqu'il interprète les critères et les applique à l'analyse de l'étude.

Les critères ici s'appliquent à toutes les études descriptives. Se référer à l'algorithme et la légende de la figure 1 «algorithme pour sélectionner le bon outil» pour être sûre que l'étude est descriptive. Ces études peuvent porter sur une exposition étudiée (p. ex. facteurs de risque, interventions ou facteurs de protection) ou un résultat étudié (p. ex. infections, maladies, comportements, effets ou conditions).

Utilisez l'outil pour évaluer les études transversales, les études écologiques, les études sur le lien épidémiologique et les rapports de cas ou séries de cas. Les études sur le lien épidémiologique comprennent les enquêtes sur les dons antérieurs, le retraçage et les contacts. Dans la présente trousse d'outils, les études transversales, écologiques et sur le lien épidémiologique sont classées dans la catégorie des études descriptives exploratoires et évaluées à l'aide du même ensemble de critères, qui figurent aux sections A et B de l'OEC d'une étude descriptive. Veuillez noter que les rapports de cas ou séries de cas font l'objet d'une section distincte dans l'OEC (parties A et C). Remplissez seulement la section qui convient au plan de l'étude que vous évaluez. Une synthèse des caractéristiques de chaque plan d'étude figure au tableau 2, de la présente trousse d'outils.

Évaluation critique de la validité d'une étude descriptive

Le principal objet d'une étude descriptive est de décrire les caractéristiques générales ou particulières d'une condition par rapport à des facteurs particuliers, comme l'exposition étudiée ou les résultats. Cela peut être utile pour découvrir des associations possibles. Dans le cadre d'une étude descriptive exploratoire, on explore souvent ces associations en plus de les décrire. Cela aide à définir les associations qui peuvent faire l'objet d'un examen plus en profondeur dans des travaux de recherche futurs à l'aide d'un plan plus rigoureux pour éprouver les hypothèses. L'évaluation critique de l'étude aide le lecteur à déterminer la validité ou la crédibilité des conclusions. Une étude descriptive est un plan de recherche faible et offre seulement des données probantes limitées. Il ne faut pas l'utiliser comme fondement d'une recommandation pratique à moins qu'aucune autre source de données probantes ne soit disponible. Les rapports de cas ou séries de cas offrent seulement des données anecdotiques qui peuvent contribuer à informer une opinion d'expert.

Consignes

Commencez l'évaluation critique en définissant le plan d'étude. Si vous ne pouvez pas définir le plan d'étude, discutez-en avec vos collègues et choisissez le plan d'étude qui s'en rapproche le plus. Classez chaque élément, à l'aide de l'outil, dans la catégorie fort, modéré ou faible, selon les critères décrits ici. Les critères ne sont pas toujours applicables au plan d'étude. À moins d'indication contraire, la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être respectés pour attribuer le classement en question. Quant aux critères d'un classement « faible », associé à l'énoncé « L'un ou l'autre des éléments suivants », l'élément devrait être classé dans la catégorie faible s'il ne possède qu'un seul critère.

Les critères à examiner sont inscrits en caractères gras italiques, et une explication plus en profondeur figure dans le dictionnaire. L'outil réserve un espace aux commentaires; l'évaluateur devrait ajouter des commentaires pour appuyer ses décisions et aider à définir les secteurs de préoccupations (p. ex. faiblesses ou limites majeures) et à déterminer s'ils ont une incidence sur le fait qu'on croit ou non les conclusions de l'auteur. Certaines études n'offrent pas assez de détails pour une évaluation adéquate de chaque critère. De telles études sont parfois non pas réellement faibles, mais simplement mal rédigées. Néanmoins, votre évaluation devrait porter uniquement sur ce qui est documenté dans le rapport d'étude et non pas sur vos hypothèses quant à l'étude. Remplissez le tableau de synthèse des données probantes durant le processus d'évaluation critique.

Si vous ne pouvez pas prendre une décision quant au classement d'un élément, écrivez un commentaire et discutez-en avec vos collègues. Les conclusions relatives à la qualité des données probantes sont généralement le fruit d'un consensus de groupe plutôt que d'une décision individuelle.

A. Filtrage des études à inclure avant l'évaluation critique

Tous les types d'études descriptives devraient être filtrés, y compris les études descriptives exploratoires et les rapports de cas ou séries de cas. Avant de prendre une décision relative à l'inclusion d'une étude, il faut la lire et brièvement définir ce qui a été fait. Si plus d'une question de recherche a été abordée ou que de multiples méthodes de recherche ont été utilisées, définissez les aspects qui se rapportent à votre question clé (voir le glossaire). Il faut noter qu'il est possible qu'un aspect d'une étude se rapporte à une question clé et qu'un autre aspect se rapporte à une différente question clé et que différentes méthodes de qualité variable soient employées. Une étude utilisée à l'appui de différentes conclusions doit être lue de nouveau en fonction de chaque question clé.

A1. L'étude a-t-elle abordé une question bien ciblée qui se rapporte à la question clé?

Une question de recherche claire devrait être abordée. La population, l'intervention et les résultats étudiés devraient être précisés. Plus la question d'étude est ciblée, plus il est probable que les auteurs pourront y répondre.

La question de recherche de l'étude devrait se rapporter à la question clé des lignes directrices. Si l'étude ne se rapporte pas à la question clé, ses résultats n'aideront probablement pas à formuler des recommandations et il est inutile de consacrer du temps à sa lecture.

Forte : Question de recherche bien ciblée, lien solide avec la question clé

Modérée : Question de recherche claire, assez bien ciblée et se rapportant à la question clé, quoiqu'elle ne permette pas toujours d'y répondre directement ou soit susceptible de donner lieu à des résultats exigeant une extrapolation pour aborder la question de recherche.

Faible : Question de recherche vague ou trop générale ou n'ayant aucun lien avec la question clé

Décision de filtrage

Tirez une conclusion quant à la poursuite de l'évaluation critique ou au rejet de l'étude. Une étude classée dans la catégorie faible devrait être évaluée avec précaution ou ne pas être évaluée du tout.

Si l'étude est descriptive exploratoire, allez à la section B, et s'il s'agit d'un rapport de cas ou de séries de cas, allez à la section C.

B. Étude descriptive exploratoire

Menez une évaluation critique de la qualité des données et de l'analyse en tenant compte du fait que les études descriptives exploratoires peuvent seulement définir des associations qui exigent un examen plus en profondeur.

B1. Évaluer si les participants étaient susceptibles de représenter la population cible

L'échantillon doit représenter la population cible pour que les inférences tirées des données soient valides pour ce groupe. On peut employer de multiples stratégies (p. ex. discussion personnelle, affiches, campagne médiatique, etc.) pour recruter ou sélectionner des participants d'une variété de sources (p. ex. hôpitaux et collectivité). L'échantillonnage aléatoire atténue le biais de sélection si tout le monde accepte de participer. ***Pour une étude transversale, une proportion élevée (> 50 %) des personnes abordées devraient accepter de participer.*** Pour les études sur le lien épidémiologique, un échantillonnage aléatoire n'est pas applicable et il est entendu que toutes les personnes exposées sont identifiées.

| | |
|------------------|--|
| Forte : | L'échantillonnage aléatoire a été employé et de multiples stratégies de recrutement ou de sélection ont été employées; recrutement d'une variété d'emplacements ou dans une variété de groupes; > 50 % ont accepté de participer. Pour les études sur le lien épidémiologique, ≥80% de ceux exposés ont été testés. |
| Modérée : | L'échantillonnage aléatoire n'a pas été utilisé, mais de multiples stratégies de recrutement ou de sélection ont été employées. Les participants ont été recrutés à partir d'une seule source susceptible d'exclure des membres de la population cible; 30-50 % ont accepté de participer. Pour les études sur le lien épidémiologique, 60-79% de ceux exposés ont été testés. |
| Faible : | Aucun échantillonnage aléatoire. Les processus de recrutement ou de sélection étaient limités. Les participants se sont proposés ou sont volontaires; <30 % ont accepté de participer. Pour les études sur le lien épidémiologique, <60% de ceux exposés ont été testés. |

B2. Évaluer les sources et les méthodes de collecte de données

Des définitions claires des facteurs étudiés devraient être appliquées uniformément à tous les participants, et les mesures devraient être aussi objectives que possible. Les données devraient être le plus exhaustives et exactes possible. Les interviewers, par exemple, peuvent poser leurs questions différemment à différentes personnes ou ne pas interpréter l'information de la même façon. Les participants ressentant des effets indésirables sur la santé sont susceptibles de ne pas se souvenir de leurs expériences de la même façon que les sujets qui n'ont pas ces effets (biais de rappel) ou les participants peuvent donner des réponses correctes sur le plan social ou politique ou dire ce qu'ils croient que le chercheur veut entendre (désirabilité sociale ou biais de déclaration). Parmi les stratégies d'atténuation de ces biais, on compte des protocoles normalisés de collecte de données, la formation des évaluateurs pour promouvoir la fidélité interévaluateurs et le respect des protocoles, la formulation des questions et la prise de mesures (p. ex. anonymat et création d'un rapport) pour mettre les participants plus à l'aise de répondre honnêtement à des questions difficiles. **De plus, l'exposition doit clairement être survenue avant le résultat (lien temporel clair).** L'incapacité d'établir clairement un lien temporel est une importante limite dans une étude transversale, mais la formulation des questions peut aider.

La source de données utilisées dans l'étude a une incidence sur sa validité et sa fiabilité et doit donc être examinée. Une source de données commune est la surveillance, mais les données peuvent aussi être tirées de sondages. Si on utilise une base de données existante, il faut déterminer chez combien de personnes on a recueilli les données et comment les participants ont été sélectionnés. Dans le cadre d'une étude écologique, on procède à la sélection et non pas au recrutement. Dans le cadre d'une étude sur le lien épidémiologique, il faut déterminer la mesure dans laquelle la base de données utilisée contenait des renseignements complets.

| | |
|------------------|---|
| Forte : | Aucune donnée manquante; les évaluateurs ont suivi une formation sur les procédures de collecte de données et les ont clairement respectées; des stratégies pour minimiser les biais associés aux procédures et mesures de collecte de données ou la formulation des questions ont été utilisées; lien temporel clair. |
| Modérée : | Les données manquantes ou inexactes sont minimales; les évaluateurs ont suivi une formation sur les procédures de collecte de données et les ont probablement respectées; des stratégies pour réduire les biais associés aux procédures et mesures de collecte de données ou la formulation des questions ont été utilisées; lien temporel clair. |
| Faible : | L'un ou l'autre des éléments suivants : Les données manquantes ou inexactes sont substantielles; on ignore si les évaluateurs ont suivi une formation sur les procédures de collecte de données et les ont suivies. On ne sait pas si les stratégies suffisaient à réduire le biais des réponses associé aux mesures de collecte de données ou à la formulation des questions; lien temporel vague. |

B3. Évaluer la validité et la fiabilité des instruments de collecte de données

Les instruments employés pour recueillir les données devraient être valides et fiables. On devrait distinguer les instruments (p. ex. guide d'entrevue, questionnaire, formulaire d'extraction des données) des méthodes de collecte de données (p. ex. entrevue, sondage, examen des dossiers). La validité tient au fait que l'instrument mesure ce qu'il est censé mesurer, et la fiabilité tient à l'uniformité du procédé. On peut employer des méthodes d'essai particulières pour évaluer la validité et la fiabilité des instruments, et on accorde une plus grande confiance à ceux ayant fait l'objet de ces essais. On peut supposer que le protocole et les essais standards de laboratoire sont valides et fiables. On peut également supposer que ces derniers ont été correctement mis en œuvre, sauf indication contraire des chercheurs.

| | |
|------------------|---|
| Forte : | Les outils sont réputés valides et fiables ou on l'a démontré. |
| Modérée : | Il n'y a eu aucune tentative visant à évaluer la validité et la fiabilité des outils, mais on peut croire à la validité du contenu du fait de la nature des questions posées et de la participation d'experts à l'élaboration des outils. |
| Faible : | Il n'y a eu aucune tentative en vue d'évaluer la validité et la fiabilité, et rien ne donne à penser que ces caractéristiques sont présentes. |

B4. Évaluer la justesse de la conduite éthique

Il faut avoir pris les mesures appropriées pour protéger les participants, surtout les groupes vulnérables, du danger, de l'exploitation et de la coercition ainsi que leur droit à l'autodétermination, à la divulgation complète, au traitement équitable et à la protection des renseignements personnels. Le consentement éclairé est l'une des grandes stratégies visant à protéger les droits. Le chercheur qui déclare que son étude a été approuvée par un conseil d'examen éthique institutionnel sans donner plus de détails doit avoir fait examiner et approuver cette déclaration au préalable. Il faut noter que, selon l'Énoncé de politiques des trois Conseils, l'évaluation de programmes et les études de surveillance n'exigent pas une approbation éthique, mais il faut tout de même prendre des mesures appropriées pour protéger les participants et leurs droits. Les investigations de santé publique, tel que les traçabilités et les investigations d'exposition ne nécessitent pas de revue éthique.

L'éthique de la recherche s'attache aussi à l'influence indue de promoteurs ou d'autres intervenants sur les méthodes ou la déclaration des résultats de façon à obtenir une conclusion favorable.

Forte : La recherche a été approuvée par un conseil d'examen éthique approprié ou on a fourni assez de détails pour indiquer que la conduite éthique était assurée. Le rapport de recherche n'a pas été influencé par un organisme de financement, un promoteur ou un conflit d'intérêts.

Modérée : Ne s'applique pas.

Faible : On n'a pas donné assez de détails pour qu'une conclusion soit tirée quant à la conduite éthique. La possibilité que le rapport de recherche ait été influencé par un organisme de financement, un promoteur ou un conflit d'intérêts n'a pas pu être exclue.

B5. Évaluer les statistiques

Le principal objet d'une étude transversale est de décrire la fréquence d'une exposition ou d'un résultat étudié (p. ex. facteur de risque, maladie, comportement). Les statistiques qui conviennent à cet exercice sont des statistiques descriptives (p. ex. moyenne ou médiane, proportion, taux) et des intervalles de confiance. On peut calculer le rapport de cotes (RC) pour évaluer les associations entre les variables et le khi carré ou le test exact de Fisher pour évaluer la signification statistique du RC. On peut recourir à la régression pour contrôler la confusion. Toutefois, il importe de noter que, puisque le plan d'étude est faible en soi, les conclusions tirées des constatations statistiquement significatives sont limitées, et le contrôle de la confusion n'est pas une grande préoccupation, car il faudra mener des travaux de recherche plus en profondeur quoi qu'il en soit.

Le principal objet d'une étude écologique est de décrire la fréquence d'une exposition en lien avec le résultat étudié (p. ex. facteur de risque, maladie, comportement). La statistique qui convient à cet égard est le coefficient de corrélation (p. ex. coefficient de corrélation de Pearson), mais le test t peut aussi être utile pour examiner les différences de moyenne. La régression est rarement utilisée pour contrôler la confusion. Veuillez noter que, comme le plan d'étude est faible en soi, les conclusions tirées des constatations statistiquement significatives sont limitées, et le contrôle de la confusion n'est pas une grande préoccupation, car il faudra mener des travaux de recherche plus en profondeur quoi qu'il en soit.

Les tests statistiques utilisés dans l'analyse devraient convenir au type et à la catégorie de données et être appliqués correctement (voir la synthèse des tests statistiques communs à l'annexe A de la trousse d'outils). Si les valeurs p et les intervalles de confiance sont utilisés, **les critères (p. ex. $\alpha=0,05$) pour déterminer la signification statistique devraient être clairs et appropriés. Il faut correctement interpréter les valeurs p** (c.-à-d. le résultat était statistiquement significatif s'il était inférieur à α , p. ex. $p < 0,05$). **Les intervalles de confiance (IC) devraient aussi être interprétés correctement lorsqu'ils sont précisés.** Les valeurs de l'IC sont toutes des valeurs réelles possibles de l'estimation ponctuelle. Si elles sont toutes supérieures à 1, ou toutes inférieures à 1, ils ont la même direction de l'intervention ou effet de l'exposition. Si certaines valeurs sont inférieures à 1 et supérieures à 1 et tous dans le même IC, les valeurs possibles de l'estimation ponctuelle pourrait également être un facteur de protection ou d'un facteur de risque, ou d'indiquer une efficacité ou une intervention inefficace. Une telle IC indique qu'on ne peut pas tirer des conclusions sur l'effet; les causes possibles incluent une puissance insuffisante, absence réelle d'effet ou de mesures inadéquates.

Dans le cadre d'une étude transversale, la puissance désigne la capacité de calculer, à partir de l'étude d'un échantillon, une estimation du facteur étudié (p. ex. prévalence) hautement susceptible de refléter la véritable estimation dans la population. La puissance d'une étude écologique désigne la capacité d'évaluer, à la lumière de l'étude, si l'association observée est uniquement le fruit du hasard. **Les études dont la taille de l'échantillon est insuffisante et qui ne sont donc pas assez puissantes feront état d'un grand IC et/ou des valeurs à l'intérieur de l'IC peut indiquer une direction différente de l'effet, ce qui limitera les conclusions tirées.** Il faut noter que la représentativité des conclusions dépend également de la représentativité de l'échantillon. **Si les coefficients de corrélation sont utilisés, les critères d'évaluation de l'ampleur de la corrélation la plus petite, moyenne et grande doivent être explicites.**

- Forte :** Les statistiques convenaient à la catégorie de données; l'IC (si fourni) était étroit, avec toute les valeurs ayant la même direction que l'intervention ou l'effet d'exposition; la puissance permettait de tirer des inférences au sujet de la population; les résultats (p. ex. moyenne, proportion, RC) ont été interprétés correctement.
- Modérée :** Les statistiques ont été employées correctement; l'IC (si fourni) était raisonnablement étroit avec une direction incertaine de l'intervention ou l'effet d'exposition; une puissance pouvant possiblement permettre de tirer des inférences au sujet de la population; les résultats (p. ex. moyenne, proportion, RC) ont été interprétés correctement.
- Faible :** L'un ou l'autre des éléments suivants : Les statistiques ne convenaient pas aux données; l'IC (si fourni) était large, et la puissance ne permettait pas de tirer des inférences au sujet de la population ou les chercheurs n'ont pas mentionné la justesse de la puissance de l'étude; les résultats (p. ex. moyenne, proportion, RC) n'ont pas été interprétés correctement.

CONCLUSION GLOBALE QUANT À L'ÉTUDE DESCRIPTIVE EXPLORATOIRE

Force du plan d'étude : Faible

Décision relative à la qualité de l'étude

La conclusion globale tirée devrait se rattacher à la qualité de l'étude et donc à la crédibilité des résultats et indiquer si toutes associations possibles observées exigent des travaux de recherche plus en profondeur.

Examinez les classements attribués aux éléments B1 à B5 à évaluer :

Classez la qualité dans la catégorie ÉLEVÉE si vous avez attribué un classement fort à la plupart des éléments à évaluer, voire tous, et qu'aucun élément n'était faible. En outre, il n'y a pas d'importants obstacles à la validité interne de l'étude ni à la capacité de tirer une conclusion quant à l'association claire entre l'exposition et le résultat étudié, entraînant de ce fait la nécessité d'un examen plus en profondeur.

Classez la qualité dans la catégorie MOYENNE si l'élément B2 ou B5 ou les deux ont été classés dans la catégorie modérée et que les autres éléments à évaluer classés comme étant faibles ou modérés ne suffisent pas à compromettre la validité interne de l'étude. Aussi, ces autres éléments ne doivent pas compromettre la capacité de tirer une conclusion selon laquelle il y a possiblement une association entre l'exposition et le résultat étudié, entraînant de ce fait la nécessité d'un examen plus en profondeur.

Classez la qualité dans la catégorie FAIBLE si les éléments B2 ou B5 ont été classés dans la catégorie faible ou si d'autres éléments classés dans cette catégorie suffisent à compromettre la capacité d'exclure d'autres explications quant aux constatations et de tirer une conclusion relative à une association possible entre l'exposition et le résultat étudié.

Décision relative au caractère direct des données probantes présentées dans l'étude

Tirez une conclusion relative au caractère direct des données probantes :

- **Les données probantes directes** proviennent des études portant précisément sur l'association étudiée.
- **L'extrapolation** est l'inférence tirée d'études qui portaient sur une question différente, mais connexe, ou la même question, mais dans des conditions artificielles (p. ex. certaines études en laboratoire). Draw a conclusion regarding the directness of evidence.

C. Rapports de cas et séries de cas

Les rapports de cas ou les séries de cas décrivent les expériences d'un patient ou de quelques patients. Les séries de cas et les rapports de cas ne contribuent pas aux données probantes, par conséquent, on ne leur attribue pas de « force du plan » au moment de l'évaluation. On peut seulement évaluer la crédibilité de la description et déterminer s'il semble y avoir des aspects qui méritent une recherche plus en profondeur.

C1. Évaluer si les participants étaient susceptibles de représenter la population cible

L'échantillon doit représenter la population cible pour que l'on puisse tirer des inférences valides pour le groupe. Comme on peut seulement tirer des conclusions limitées d'une série de cas, la représentativité n'est pas une préoccupation importante. Si les séries de cas indiquent qu'une recherche approfondie est justifiée, la représentativité peut être abordée dans des études ultérieures.

| | |
|------------------|---|
| Forte : | Les participants possédaient des caractéristiques semblables à celles d'un groupe étudié plus grand. |
| Modérée : | Ne s'applique pas. |
| Faible : | Les participants n'avaient pas de caractéristiques semblables à celles d'un plus grand groupe étudié. |

C2. Évaluer la crédibilité de la description

Lorsqu'on évalue la crédibilité de la description, il faut examiner la validité et la fiabilité des sources de données et se pencher sur l'objectivité des méthodes de collecte de données et le caractère exhaustif et exact des détails.

| | |
|------------------|---|
| Forte : | Les méthodes de collecte de données étaient objectives, les détails sont complets et il manque peu d'information, voire pas du tout, et on a fait des efforts pour atténuer le biais d'information. |
| Modérée : | Ne s'applique pas. |
| Faible : | Les méthodes de collecte de données n'étaient pas objectives ou les détails sont incomplets et on a déployé peu d'efforts pour atténuer le biais d'information. |

CONCLUSION GLOBALE QUANT AUX RAPPORTS DE CAS OU AUX SÉRIES DE CAS

La conclusion globale devrait traiter de la crédibilité du rapport et déterminer s'il y a des aspects qui méritent une recherche plus en profondeur.

OUTIL D'ÉVALUATION CRITIQUE D'UNE ÉTUDE DESCRIPTIVE

Question clé : _____

Auteur : _____ Année : _____ Numéro de référence : _____

Titre : _____

Évaluateur : _____ Date : _____

Voir le dictionnaire de l'outil d'évaluation critique d'une étude descriptive pour obtenir la liste complète des critères. Seulement remplir la section qui convient au plan d'étude évalué.

À moins d'indication contraire (par l'énoncé « l'un ou l'autre des éléments suivants »), la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être présents pour que l'on attribue à l'élément le classement en question.

| A. Question de filtrage | | | |
|----------------------------------|--|---|--|
| | Forte | Modérée | Faible |
| A1. Question de recherche | Bien ciblée. Lien solide avec la question clé. <input type="checkbox"/> | Assez bien ciblée. Lien avec la question clé. <input type="checkbox"/> | Vague ou trop générale. Aucun lien avec la question clé. <input type="checkbox"/> |
| Commentaires : | | | |

| Décision relative au filtrage | |
|--|--|
| <input type="checkbox"/> Rejeter (si faible) | OU <input type="checkbox"/> Poursuivre |

| B. Étude descriptive exploratoire | | | |
|--|--|---|--|
| | Forte | Modérée | Faible |
| B1. Participants de l'étude représentatifs de la population cible | Échantillonnage aléatoire ou multiples méthodes de recrutement et de sélection à partir de différents endroits ou dans différents groupes; > 50 % ont accepté de participer (ou ≥80% des exposés sont testés). <input type="checkbox"/> | Aucun échantillonnage aléatoire, mais de multiples stratégies de recrutement ou de sélection ont été employées. Source de participants unique; 30-50 % ont accepté de participer (ou 60-79% des exposés sont testés). <input type="checkbox"/> | Aucun échantillonnage aléatoire. Processus de recrutement ou de sélection limités. Les participants étaient des volontaires; < 30 % ont accepté de participer (ou < 60% des exposés sont testés). <input type="checkbox"/> |
| B2. Sources et méthodes de collecte de données | Aucune donnée manquante. Évaluateurs formés et ont clairement respectés les procédures. Biais associés aux procédures et mesures de collecte de données sont minimisés. Lien temporel clair. <input type="checkbox"/> | Les données manquantes ou inexactes sont minimales. Évaluateurs formés et ont probablement respectés les procédures. Biais associés aux procédures et mesures de collecte de données sont réduits. Lien temporel clair. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Les données manquantes ou inexactes sont substantielles. On ignore si les évaluateurs ont suivi une formation. On ne sait pas si le biais a été réduit. Lien temporel vague. <input type="checkbox"/> |

| | | | |
|---|---|---|--|
| B3. Instruments de collecte de données | Outils réputés valides et fiables. <input type="checkbox"/> | Aucune tentative visant à évaluer la validité et la fiabilité des outils. On peut croire à la validité à la lumière des questions posées et de l'expertise des chercheurs. <input type="checkbox"/> | Aucune tentative d'évaluer la validité et la fiabilité des outils; rien ne donne à penser que c'est le cas. <input type="checkbox"/> |
| B4. Éthique <input type="checkbox"/> Ne s'applique pas (voir dictionnaire) | Approuvée par un conseil d'examen éthique approprié ou le contenu révèle qu'une conduite éthique a été assurée. Le rapport de recherche n'a pas été influencé. <input type="checkbox"/> | Ne s'applique pas. | Détails fournis ne suffisant pas à établir la conduite éthique. La probabilité que le rapport de recherche a été influencé n'a pu être exclue. <input type="checkbox"/> |
| B5. Statistiques (voir tableau 5) Évaluer l'IC si fourni | Statistiques appropriées utilisées (descriptives). IC étroit avec toutes les valeurs ayant la même direction de l'effet. Une puissance clairement adéquate. Résultats interprétés correctement. <input type="checkbox"/> | Statistiques appropriées employées. IC raisonnablement étroit ayant une direction incertaine de l'effet. Une puissance possiblement adéquate. Résultats interprétés correctement. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Les statistiques ne convenaient pas aux données. IC large. Une puissance inadéquate. Résultats n'ont pas été interprétés correctement. <input type="checkbox"/> |

| Conclusion globale | |
|--|--|
| <p>a) *Force du plan d'étude : Faible</p> <p>b) Décision relative à la qualité de l'étude Examinez les classements que vous avez attribués aux éléments B1 à B5 à évaluer et définissez le classement approprié pour la qualité.</p> <p style="text-align: center;"> <input type="checkbox"/> Élevée <input type="checkbox"/> Moyenne <input type="checkbox"/> Faible </p> <p>Classez la qualité dans la catégorie ÉLEVÉE si vous avez attribué un classement fort à la plupart des éléments à évaluer, voire tous, et qu'aucun élément n'était faible. En plus, il n'y a pas d'importants obstacles à la validité interne de l'étude ni à la capacité de tirer une conclusion quant à l'association claire entre l'exposition et le résultat étudié.</p> <p>Classez la qualité dans la catégorie MOYENNE si les éléments B2 ou B5 ont été classés au moins dans la catégorie modérée et que les autres éléments à évaluer classés comme étant faibles ou modérés ne suffisent pas à compromettre la validité interne de l'étude. Aussi, ces autres éléments ne doivent pas compromettre la capacité de tirer une conclusion selon laquelle il y a probablement une association entre l'exposition et le résultat étudié.</p> <p>Classez la qualité dans la catégorie FAIBLE si les éléments B2 ou B5 ont été classés dans la catégorie faible ou si d'autres éléments classés dans cette catégorie suffisent à compromettre la capacité d'exclure d'autres explications quant aux constatations et de tirer une conclusion relative à l'association entre l'exposition et le résultat étudié.</p> <p>c) Décision relative au caractère direct des données probantes</p> <p style="text-align: center;"> <input type="checkbox"/> Direct <input type="checkbox"/> Extrapolation </p> | |
| <p>Commentaires :</p> | |

| C. Séries de cas et rapports de cas | | | |
|--|--|---|--|
| | Forte | Modérée | Faible |
| C1. Participants de l'étude représentatifs de la population cible | Les participants possédaient des caractéristiques semblables à celles du plus grand groupe étudié. <input type="checkbox"/> | Ne s'applique pas <input type="checkbox"/> | Les participants n'avaient pas de caractéristiques semblables à celles du plus grand groupe étudié. <input type="checkbox"/> |
| C2. Qualité de la description | Les méthodes de collecte de données étaient objectives. Biais d'information réduit. Quantité minimale de données manquantes. <input type="checkbox"/> | Ne s'applique pas <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Les méthodes de collecte de données n'étaient pas objectives ou les détails sont incomplets. Peu d'efforts utilisés pour réduire le biais d'information. <input type="checkbox"/> |
| Commentaires : | | | |

DICTIONNAIRE DE L'OUTIL D'ÉVALUATION CRITIQUE D'UNE REVUE DE LITTÉRATURE

Introduction

L'outil d'évaluation critique (OEC) d'une revue de littérature a pour objet d'aider à l'évaluation de l'utilité des résultats d'une revue de littérature publiée. L'objet du présent dictionnaire consiste à fournir des renseignements de base concernant les types de revues de littérature existants et à décrire les éléments de l'OEC pour aider les évaluateurs à classer la qualité de la revue. Il importe de comprendre que le dictionnaire ne présente pas une explication en profondeur de tous les concepts et ne les illustre pas à l'aide de tous les exemples possibles.

Types de revues

Il y a différents types d'articles de revue de littérature, et les lecteurs devraient connaître les distinctions. La qualité de la revue dépend des auteurs : leur capacité de trouver toutes les études pertinentes, leurs compétences d'évaluation critique et leur capacité de synthétiser et de communiquer les conclusions pertinentes. Pour les besoins de l'évaluation critique, nous avons défini trois catégories de revues de littérature.

La **revue narrative** synthétise les renseignements liés au sujet, mais ne donne que des sommaires des résultats. Elle contient des références, mais la description et l'évaluation critique des études individuelles sont limitées. Si on ne traite pas des méthodes de recherche et d'évaluation critique, le lecteur doit se fier aux compétences des auteurs et ne peut pas juger lui-même la qualité de la méthode employée dans l'évaluation. Les revues narratives sont utiles pour trouver des références d'études individuelles pouvant faire l'objet d'une évaluation critique, mais elles ne conviennent pas à l'élaboration de lignes directrices. Il en est ainsi parce que la critique d'études individuelles est essentielle à la prise de décisions quant à la qualité des données probantes, et ces revues n'offrent pas de renseignements à ce sujet.

La **revue systématique de littérature** suit et décrit un protocole structuré pour trouver toutes les études éligibles sur un sujet, y compris celles qui n'ont pas été publiées ou dont la langue de publication n'est pas l'anglais, et en faire une évaluation critique. Les revues systématiques de littérature sont ainsi plus rigoureux que les autres types de revues. Une limite de la revue systématique de littérature tient au fait que, habituellement, il ne porte que sur les études d'intervention fortes, alors que peu de ces études peuvent exister sur un sujet donné. Il faut tenir compte de la méthode de sélection et d'évaluation des articles des auteurs pour déterminer si leurs conclusions devraient être acceptées. Sinon, les conclusions de la revue ne doivent pas être acceptées et les études primaires doivent être évaluées individuellement. Il importe de noter que le présent outil d'évaluation critique n'est pertinent qu'aux revues portant sur des études quantitatives, et non pas des études appuyées sur une méthode de recherche qualitative.

On considère que la revue systématique Cochrane et les lignes directrices fondées sur des données probantes publiées sont des revues systématiques de littérature. Il est important de noter que des lignes directrices fondées sur des données probantes peuvent être traitées comme une revue narrative, auquel cas l'accent est mis sur les recommandations, mais elles s'appuient en vérité sur un processus d'examen exhaustif et systématique. Il faut tenir compte de la méthode d'examen avant de décider de l'emploi des recommandations.

La méta-analyse est une revue systématique de littérature qui repose en outre sur la mise en commun de données quantitatives tirées des études primaires et la nouvelle analyse de ces données à l'aide de méthodes statistiques établies. Les données de certaines études peuvent se perdre parce qu'elles ont été refondues en catégories ou que l'on a créé des définitions qui s'appliquent à toutes les études comprises, mais la mise en commun des données fait grossir l'échantillon et fait augmenter la puissance qui permet de tirer des conclusions statistiquement significatives. Il n'est pas toujours possible de procéder à la méta-analyse si les données tirées des études primaires sont trop différentes.

Approche d'un rapport de revue de littérature

La revue de littérature peut être complexe, parce qu'elle repose sur plusieurs études et par conséquent exige une approche différente de celle des autres types d'études. Il est utile de commencer par lire le résumé pour obtenir un aperçu de l'objet du rapport et de ses principaux résultats. Toutefois, il importe de lire les sections relatives aux méthodes et aux résultats, car le résumé n'est pas toujours complet ni exact.

La plupart des rapports reposent sur une structure semblable et commencent par donner un bref contexte et une explication du problème; on définit ensuite les objectifs. On décrit les méthodes; la description est habituellement brève dans une revue narrative, détaillée dans une revue systématique de littérature et varie dans d'autres rapports. Les résultats peuvent être regroupés par thème ou par étude ou les deux. Le rapport se termine habituellement par des conclusions et des recommandations. Les structures sont semblables dans tous les rapports, mais la quantité de détails varie et il est possible que les titres des différentes sections soient différents. Il peut être utile de jeter un coup d'œil rapide sur l'ensemble de l'article avant d'en entreprendre la lecture pour définir la structure et les grands titres du rapport, ce qui aide à trouver la section qui doit être étudiée pour répondre aux questions soulevées dans l'OEC.

À l'exception des méta-analyses, qui reposent sur des plans solides, la force du plan d'une étude ne s'applique pas aux revues de littérature. Il en est ainsi parce que les revues de littérature sont non pas des études de recherche primaires, mais une synthèse des conclusions de plusieurs études de recherche.

Consignes

Commencez l'évaluation critique en définissant le plan d'étude. Si vous ne pouvez pas définir le plan d'étude, discutez-en avec vos collègues et choisissez le plan d'étude qui s'en rapproche le plus. Classez chaque élément, à l'aide de l'outil, dans la catégorie fort, modéré ou faible, selon les critères décrits ici. Les critères ne sont pas toujours applicables au plan d'étude. À moins d'indication contraire, la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être respectés pour attribuer le classement en question. Quant aux critères d'un classement « faible », associé à l'énoncé « l'un ou l'autre des éléments suivants », l'élément devrait être classé dans la catégorie faible s'il ne possède qu'un seul critère.

Les critères à examiner sont inscrits en caractères gras italiques, et on donne une explication plus en profondeur. Certaines études n'offrent pas assez de détails pour une évaluation adéquate de chaque critère. De telles études sont parfois non pas réellement faibles, mais simplement mal rédigées. Néanmoins, votre évaluation devrait porter uniquement sur ce qui est documenté dans le rapport d'étude et non pas sur vos hypothèses quant à l'étude.

L'outil réserve un espace aux commentaires; l'évaluateur devrait ajouter des commentaires pour appuyer ses décisions et aider à définir les secteurs de préoccupations (p. ex. faiblesses ou limites majeures) et à déterminer s'ils contribueraient à une conclusion erronée au sujet de l'existence d'une association ou de sa force (surestimation ou sous-estimation d'un effet).

Si vous ne pouvez pas prendre une décision quant au classement d'un élément, écrivez un commentaire et discutez-en avec vos collègues. Les conclusions relatives à la qualité des données probantes sont généralement le fruit d'un consensus de groupe plutôt que d'une décision individuelle.

Filtrage des revues de littérature à inclure avant l'évaluation critique

1. La revue permettait-elle de répondre à une question bien ciblée se rapportant à la question clé?

Une question centrale claire devrait être abordée. La population, l'intervention et les résultats étudiés devraient être précisés. Plus la question de la revue est ciblée, plus les auteurs sont susceptibles de pouvoir y répondre. Les questions trop générales (p. ex. comparer un trop grand nombre d'interventions ou inclure trop de groupes cibles) sont plus difficiles à moins qu'on ne s'attende à obtenir le même effet dans le cas d'un éventail de patients, d'interventions ou de résultats.

La question centrale de l'analyse devrait se rapporter à la question clé. Si l'analyse ne se rapporte pas à la question clé, ses résultats sont peu susceptibles de contribuer à la formulation de recommandations, et il n'est pas très utile de consacrer du temps à sa lecture.

Forte : Question bien ciblée, lien solide avec la question clé.

Modérée : Question claire, assez bien ciblée et se rapportant à la question clé, quoiqu'elle ne permette pas toujours d'y répondre directement.

Faible : Question centrale vague ou trop générale ou n'ayant aucun lien avec la question clé.

2. La méthode de la revue est-elle acceptable sur le plan du nombre d'études incluses et de l'évaluation critique de ces études?

Les études incluses dans la revue doivent se rattacher à la réponse à la question définie. Si une étude ne se rapporte pas à la question définie, alors les résultats ne contribueront pas à la formulation de recommandations.

Les critères d'inclusion et d'exclusion devraient être explicites, clairs et raisonnables par rapport à la population, à l'intervention, au résultat et au plan d'étude. Si on ignore comment ont été prises les décisions menant à l'inclusion ou à l'exclusion d'une étude, alors il est difficile de déterminer s'il manque des études appropriées.

Chaque étude doit avoir fait l'objet d'une évaluation critique de façon systématique et uniforme à l'aide des critères reconnus pour mesurer la qualité de la méthode.

L'application des critères reconnus de façon uniforme permet de s'assurer que chaque étude a fait l'objet d'une évaluation appropriée.

Les résultats de la critique de chaque étude devraient être clairs pour que les lecteurs comprennent les forces et les limites de chaque étude et ses résultats. La revue ne doit pas être limitée à la présentation des résultats d'études individuelles. Le lecteur devrait pouvoir déterminer lui-même si les auteurs ont défini toutes les forces et les limites clés, de façon à ce qu'il puisse décider de croire ou non les conclusions des auteurs.

Les études examinées devraient comprendre des études analytiques (si disponible) et pas seulement des études descriptives. Une étude descriptive ne présente pas assez de données probantes scientifiques quant à l'existence ou à la force d'un lien entre l'exposition et le résultat. Les recommandations liées aux changements de pratiques devraient reposer sur des données probantes de qualité. Les meilleures données probantes sont tirées d'essais contrôlés avec contrôle rigoureux des autres explications. Comme on n'aura pas toujours mené des essais contrôlés, certaines conclusions devront être fondées sur les données probantes tirées d'études par observation (cas-témoin ou de cohorte) qui ont été menées de façon rigoureuse.

- Forte :** Les études se rapportant à la question clé incluses. Inclusion d'études analytiques. Critères d'inclusion clairs définis. Les études ont été évaluées de façon systématique et uniforme à l'aide de critères reconnus et démontrent des résultats clairs.
- Modérée :** Les études se rapportant à la question clé incluses. Études analytiques incluses. Les critères d'inclusion ou les critères liés à l'évaluation critique n'ont pas nécessairement été clairement définis. Les résultats de la critique de chaque étude étaient clairs.
- Faible :** L'un ou l'autre des éléments suivants : Les études se rapportant à la question clé n'ont pas été incluses, aucune étude analytique incluse, les critères d'inclusion ne sont pas clairement définis, on n'a pas donné les résultats de l'évaluation critique de chaque étude.

DÉCISION RELATIVE AU FILTRAGE

Déterminez s'il convient de poursuivre l'évaluation critique ou de rejeter la revue de littérature.

Si la revue obtient un classement faible pour l'élément à évaluer n° 2, interrompez l'évaluation critique et ne remplissez pas le tableau de synthèse des données probantes pour l'ensemble de la revue de littérature. Choisissez plutôt des études qui se rapportent à la question clé et évaluez chaque étude individuelle à l'aide de l'outil d'évaluation critique d'une étude analytique.

Si l'élément n° 2 est modéré ou fort, mais que l'élément n° 1 est faible, déterminez soigneusement s'il y a intérêt à poursuivre.

Évaluation de la méthode

3. Les auteurs ont-ils mené des travaux de recherche complets portant sur des articles pertinents?

Les auteurs devraient effectuer des recherches complètes dans plusieurs bases de données à l'aide de plusieurs termes de recherche. Des recherches documentaires en profondeur offrent un éventail complet et équilibré (qui reflète à la fois les résultats positifs et négatifs) des études à inclure. La recherche doit aller au-delà de MEDLINE, car cette base de données ne représente qu'une fraction du nombre total de revues publiées à l'échelle mondiale. En raison du manque d'uniformité de l'indexation des articles dans les bases de données, il faudra inclure plusieurs termes de recherche dans une stratégie pour s'assurer que la majorité des articles pertinents ont été trouvés.

Les évaluateurs devraient rechercher des études non rédigées en anglais. Il faut rechercher des études utiles dans d'autres langues et les inclure. **Les évaluateurs devraient obtenir des articles pertinents qui ont été mentionnés dans d'autres études.** Les recherches devraient comprendre un examen des bibliographies des études incluses pour trouver des références supplémentaires susceptibles de ne pas avoir été détectées dans le cadre de la recherche dans les bases de données.

Les évaluateurs devraient chercher la littérature grise (p. ex. rapports gouvernementaux) et des études non publiées. La recherche devrait comprendre des études non publiées, en raison du biais de publication (on publie un plus grand nombre d'études démontrant des résultats favorables). Il est possible de le faire en consultant directement les chercheurs. Il est aussi possible de trouver des chercheurs et des études non publiées dans des résumés, dans le cadre de conférences, auprès d'associations professionnelles ou d'entreprises pharmaceutiques, etc.

Forte : Recherche complète dans plusieurs bases de données et sources, les bibliographies, la littérature rédigée dans une langue autre que l'anglais et la littérature grise ou non publiée.

Modérée : Recherche complète dans les bases de données, y compris la littérature rédigée dans une langue autre que l'anglais. On peut ou non avoir consulté les bibliographies ou recherché de la littérature grise ou non publiée.

Faible : Recherche limitée sur le plan du nombre de bases de données et de la littérature rédigée dans une langue autre que l'anglais. Absence de recherche de la littérature grise ou non publiée.

4. À quel point le processus de revue était-il rigoureux?

Il faut seulement inclure les études qui respectent les critères à cet effet selon l'évaluation indépendante de plus d'un évaluateur. La revue devrait exclure toutes les études prélevées dans le cadre de la recherche qui ne respectaient pas les critères d'inclusion prédéterminés, p. ex. celles qui n'abordaient pas la question clé ou ne reposaient pas sur un plan d'étude acceptable.

Plus d'un évaluateur devrait examiner chaque étude à l'aide des mêmes critères pour ensuite en arriver à une bonne entente sur les résultats de l'évaluation critique. Comme l'évaluation critique risque de devenir un processus quelque peu subjectif, il est souhaitable que deux évaluateurs indépendants évaluent les articles à l'aide des mêmes critères et tirent des conclusions semblables.

Dans la cadre d'une méta-analyse, les données devraient être prélevées indépendamment par deux évaluateurs à l'aide d'une formule normalisée créée précisément pour l'étude et éprouvée. Les différences dans les données prélevées devraient faire l'objet d'une analyse et être résolues à la suite d'un consensus.

Forte : Seules les études qui respectent les critères d'inclusion ont été incluses, et plus d'un évaluateur a filtré et a mené une évaluation critique de chaque étude à l'aide des mêmes critères pour ensuite en arriver à une bonne entente.

Modérée : Les critères d'inclusion et d'évaluation critique ont été appliqués, mais il n'y a pas eu plus d'un évaluateur ou on ignore quels critères ont été utilisés.

Faible : Aucun critère d'inclusion et d'évaluation critique utilisée (ou on l'ignore).

5. Si les résultats ont été combinés dans le cadre d'une méta-analyse, était-il raisonnable de procéder ainsi?

Les résultats de chaque étude individuelle devraient être inclus de sorte que le lecteur puisse déterminer lui-même si les résultats combinés représentent bien le portrait réel démontré par le résultat individuel. **On devrait faire un essai pour s'assurer que les études peuvent être combinées (p. ex. khi carré pour définir l'homogénéité). Les études combinées ne devraient pas être très différentes du point de vue de la population, des interventions, des comparaisons établies ou des résultats mesurés.**

Une hétérogénéité significative n'est pas désirable, et une évaluation visuelle du schéma en forêt établissant le degré de variation entre les résultats des études individuelles devrait démontrer une certaine uniformité sur le plan de l'orientation des résultats (homogénéité). **En cas d'hétérogénéité, on doit utiliser un modèle à effets aléatoires et les auteurs devraient aborder le caractère approprié sur le plan clinique de la combinaison des études ainsi que les sources possibles d'hétérogénéité.**

L'estimation mise en commun devrait correctement être interprétée à l'aide de statistiques sommaires appropriées (p. ex. rapport de cotes, risque relatif, etc.) en fonction du type de données (voir la « synthèse des tests statistiques communs » à l'annexe A).

- Forte :** Les études combinées ne variaient pas beaucoup sur le plan de la population, de l'intervention appliquée et des résultats mesurés. Une hétérogénéité minimale existe entre les résultats des études individuelles. Statistiques sommaires appropriées employées.
- Modérée :** Les études combinées ne variaient pas beaucoup sur le plan de la population, de l'intervention appliquée et des résultats mesurés. Il existe une hétérogénéité significative, mais les auteurs ont correctement abordé ce sujet. Les statistiques sommaires signalées semblent raisonnables.
- Faible :** Les études combinées variaient considérablement sur le plan de la population, de l'intervention appliquée ou des résultats mesurés. Il existe une hétérogénéité significative, et on ne l'a pas mentionnée. Les statistiques sommaires ne semblent pas raisonnables.

DÉCISION RELATIVE À LA MÉTHODE

Tirez une conclusion quant à la force de la méthode de la revue.

Méthode de la revue faible :

Si l'élément n° 4 est faible, cessez d'évaluer la revue de littérature et n'ajoutez pas les données dans le tableau de synthèse.

Si l'élément n° 3 et/ou l'élément n° 5 est faible, déterminez soigneusement quel est l'intérêt de poursuivre.

Si l'évaluation critique est interrompue, définissez les études de la revue de littérature qui se rapportent à la question clé et évaluez-les individuellement à l'aide de l'outil d'évaluation critique approprié.

Méthode d'analyse modérée ou forte :

Si les éléments 3 à 5 sont modérés ou forts, poursuivez l'évaluation.

Évaluation des résultats de l'étude

6. Les résultats ont-ils clairement été décrits et interprétés de façon significative?

Remarque : Aux fins de l'évaluation critique d'une méta-analyse, sautez cet élément et évaluez les résultats à l'aide de l'élément n° 7.

La signification statistique du résultat devrait être interprétée correctement. Les critères établissant la signification devraient être clairs, et les résultats devraient être interprétés selon des normes reconnues. Si les intervalles de confiance ont été déclarés, la décision d'utiliser ou non l'intervention devrait être la même, peu importe où se trouve la véritable valeur (p. ex. dans un contexte clinique) sur le spectre de la limite de confiance du résultat.

L'ampleur de l'effet devrait être significative sur le plan clinique. Cet élément est soumis au jugement et varie selon le sujet et le contexte.

Puisqu'il n'y a pas eu de méta-analyse, les résultats tirés des études devraient être décrits en fonction de leur similitude ou de leur différence. Comme il n'y a pas de mesure sommaire unique dans les circonstances, il est utile d'avoir une description de l'éventail des résultats et de toute tendance démontrée (p. ex. si les résultats sont en conflit, si l'effet est toujours positif ou négatif ou fort ou variable). **On offre une explication raisonnable de toute variation des résultats.**

Forte : Interprétation correcte de la signification statistique et de l'intervalle de confiance ou synthèse raisonnable de la tendance et interprétation raisonnable de l'effet potentiel sur les patients (p. ex. signification clinique, raison de la variation).

Modérée : Interprétation correcte de la signification statistique et de l'intervalle de confiance ou synthèse raisonnable de la tendance, mais aucune analyse de la signification clinique ou de la raison de la variation.

Faible : N'a pas correctement interprété les résultats.

7. Seulement pour les méta-analyses : Évaluer l'ampleur et la précision de l'effet du traitement

Parmi les méthodes couramment utilisées pour déclarer les résultats d'une méta-analyse, on compte le rapport de cotes ou le risque relatif (si le résultat présente une dichotomie, p. ex. maladie et absence de maladie) et les différences de moyenne (si le résultat est continu, p. ex. mesure de la tension artérielle). **Un risque relatif supérieur à 1 indique que le résultat énoncé est plus grand dans le groupe exposé que dans le groupe non exposé** (c.-à-d. risque augmenté), **tandis qu'un risque relatif inférieur à 1 indique que le résultat énoncé est plus bas dans le groupe exposé que dans le groupe non exposé** (c.-à-d. risque réduit). **Un ratio de un révèle l'absence de différences dans le résultat** (c.-à-d. le risque est probablement le même dans le groupe exposé et le groupe non exposé). **Un rapport de cotes est une estimation du risque relatif et est interprété de la même façon en terme de risque, même s'il évalue les cotes d'exposition de ceux qui ont le résultat et non le risque du résultat de ceux qui ont été exposés.**

Le rapport de cotes et le risque relatif énoncés dans une méta-analyse devraient être accompagnés d'intervalles de confiance (IC). **La largeur de l'intervalle de confiance (IC) témoigne de la précision de l'estimation. Plus l'intervalle de confiance est large, moins l'estimation est précise. La décision d'utiliser ou non l'intervention (p. ex. dans un contexte clinique) devrait être la même, peu importe où s'inscrit la véritable valeur dans le spectre de la limite de confiance du résultat.**

Le schéma en forêt devrait démontrer de petites différences sur le plan de l'ampleur de l'effet du traitement entre les études, et les IC des différentes estimations ponctuelles des études devraient largement se chevaucher. On accorde plus de poids aux résultats d'études plus vastes qui donnent plus de renseignements, car ils sont susceptibles de se rapprocher davantage du véritable effet. Un effet de traitement global est calculé comme une moyenne pondérée des statistiques sommaires individuelles. **Il faut déclarer un IC de 95 % ou de 99 %.**

Une méta-analyse qui n'est pas assez puissante démontre généralement que l'effet du traitement est extrêmement bénéfique. Dans de tels cas, il faut entreprendre une évaluation minutieuse. Le nombre total de participants ou d'interventions mis en commun (taille de l'échantillon) a une plus grande incidence sur la puissance de l'étude que le nombre total d'études mises en commun, ce qui signifie que la mise en commun d'études primaires vastes est bénéfique.

Forte : On signale des intervalles de confiance (IC) de 95 % ou 99 %. La différence de l'ampleur de l'effet du traitement entre les études individuelles était minime et les IC se chevauchaient largement. La puissance semble suffisante. Interprétation correcte de la signification statistique et de l'intervalle de confiance.

Modérée : On a signalé des IC de 95 % ou de 99 %. Il y avait une certaine différence dans l'ampleur de l'effet du traitement entre les études individuelles, et les IC se chevauchaient quelque peu. La puissance semblait suffisante. Interprétation correcte de la signification statistique et de l'intervalle de confiance.

Faible : L'un ou l'autre des éléments suivants (même si on signale un IC global de 95 % ou 99 %) : La différence de l'ampleur de l'effet du traitement entre les études individuelles était importante et les IC se chevauchaient peu, voire ou pas du tout. Puissance insuffisante. Résultats mal interprétés.

DÉCISION RELATIVE AUX RÉSULTATS

Faites la synthèse des résultats en fonction des questions suivantes et ajoutez-la dans votre tableau de synthèse des données probantes.

1. Y a-t-il un effet clair?
2. Les résultats sont-ils uniformes d'une étude à l'autre?
3. Le nombre d'études ayant contribué à la décision relative à l'effet clair est-il suffisant (au moins quatre)?
4. Les données probantes sont-elles directes?
5. L'effet est-il significatif sur le plan clinique?
6. Dans le cas d'une méta-analyse, les données ont-elles correctement été mises en commun et a-t-on mené une analyse statistique adéquate?

Si la réponse à chaque question est oui, alors il peut être nécessaire d'évaluer l'applicabilité à l'aide des éléments à évaluer n° 8 et 9 ci-dessous.

Si la réponse à l'une de ces questions est négative, alors n'évaluez pas l'applicabilité; passez à la question 10 et tirez une conclusion globale, mais n'énoncez pas de recommandation.

DÉCISION RELATIVE AU CARACTÈRE DIRECT DES DONNÉES PROBANTES PRÉSENTÉES DANS L'ÉTUDE

Tirez une conclusion relative au caractère direct des données probantes :

- **Les données probantes** directes proviennent des études portant précisément sur l'association étudiée.
- **L'extrapolation** est l'inférence tirée d'études qui portaient sur une question différente, mais connexe, ou la même question, mais dans des conditions artificielles.

Évaluation de l'applicabilité

8. Les résultats peuvent-ils être appliqués à la population étudiée (utilisateurs potentiels des lignes directrices)?

L'échantillon de population ou le contexte abordé par l'analyse devrait être semblable aux groupes ou au contexte étudiés pour que l'on puisse appliquer l'intervention.

L'inclusion de différentes études fait qu'il est beaucoup plus probable que les résultats s'appliquent à un vaste pan de la population étudiée. Il faut déterminer si les différences contribueront à l'application ou la compromettront.

Forte : Caractéristiques de l'échantillon (population et contexte) très semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

Modérée : Caractéristiques de l'échantillon quelque peu semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

Faible : Caractéristiques de l'échantillon pas du tout semblables à celles du groupe auprès duquel on souhaite généraliser les résultats.

9. A-t-on examiné tous les résultats importants?

Une quantité suffisante de renseignements devrait être présentée au sujet des effets indésirables et des coûts ou des résultats étudiés par d'autres intervenants. D'autres résultats peuvent compromettre l'applicabilité d'une intervention, et on doit en tenir compte lorsqu'on prend une décision relative à l'application potentielle de l'intervention. Les intervenants dont les points de vue doivent être pris en considération comprennent les patients, les familles, les fournisseurs de soins, les décideurs, les professionnels et la collectivité.

| | |
|------------------|---|
| Forte : | L'intervention est très susceptible d'être facilement mise en œuvre dans d'autres contextes. |
| Modérée : | L'intervention est quelque peu susceptible d'être facilement mise en œuvre dans d'autres contextes. |
| Faible : | L'intervention est peu susceptible d'être facilement mise en œuvre dans d'autres contextes. |

Conclusion globale (pour toutes les revues, dont les méta-analyses)

10. Quelle conclusion peut-on tirer à la lumière des données probantes contenues dans la revue?

Remarque : Si l'évaluation critique de la revue de littérature n'a pas été menée parce qu'elle a été rejetée à l'étape du filtrage ou de l'évaluation de la méthode ou que les données probantes de l'article ne suffisent pas à établir une recommandation, ceci devrait être précisé dans l'outil et ne pas tirer d'autres conclusions ultérieures.

Si l'étude n'a pas été rejetée à l'étape du filtrage ou en raison de méthodes de revue faibles et que les données probantes suffisent à établir une recommandation, alors il faut tirer une conclusion finale à la lumière des résultats de la revue. Une revue systématique de littérature de haute qualité offre une bonne évaluation des données probantes au moment de la rédaction de la revue. Les résultats doivent être considérés dans le contexte des études qui auraient été menées depuis le moment de la recherche documentaire effectuée par les évaluateurs.

Il importe de noter que les conclusions varieront en fonction de la qualité des méthodes de la revue ainsi que des véritables résultats des études incluses. Remplissez le tableau de synthèse des données probantes et tirez des conclusions globales à l'aide des lignes directrices qui suivent.

CONCLUSION GLOBALE

Force du plan d'étude (applicable aux méta-analyses seulement) : Forte

Revue systématique de littérature et revue narrative : Aucun classement

Décision relative à la qualité de l'étude

La conclusion globale devrait être relative à la qualité (rigueur) des méthodes de la revue ainsi qu'à la qualité des études de recherche incluses dans la revue de littérature et, par conséquent, à la crédibilité du corpus de données probantes abordées dans la revue de littérature. Avant de formuler une recommandation fondée sur la revue de littérature, il faut déterminer s'il existait une association claire entre l'exposition et le résultat et si les échantillons des études incluses dans la revue de littérature ressemblent au groupe auprès duquel on souhaite généraliser les résultats.

Examinez les classements attribués à la méthode et aux décisions relatives aux résultats :

Classez la qualité dans la catégorie ÉLEVÉE si la décision relative à la méthode était forte et que la conclusion globale par rapport à l'association entre l'exposition et le résultat étudié provient de quatre études, ou plus, ayant une force de plan d'étude forte et dont la qualité est élevée.

Classez la qualité dans la catégorie MOYENNE si les méthodes de revue étaient classées dans la catégorie modérée ou que les méthodes ont été classées dans la catégorie forte, mais que moins de quatre études ont contribué à la conclusion globale ou que la force des plans d'études incluses n'était pas forte et que leur qualité n'était pas élevée.

Toute revue de littérature ayant une méthode faible devrait être considérée comme de faible qualité et aurait dû être rejetée du processus d'évaluation.

REEMPLIR LE TABLEAU DES DONNÉES PROBANTES

Lorsque les données probantes étaient suffisantes à la formulation d'une recommandation et que les résultats étaient applicables à la population étudiée :

Synthétisez ce qui suit et ajoutez le résultat au tableau de synthèse des données probantes (à l'aide des définitions pour évaluer les données probantes) :

- a) conclusion globale et résultats relatifs à l'effet;
- b) nombre d'études pour chaque force du plan et qualité de la revue systématique de littérature (p. ex. cinq études ayant une force de plan d'étude forte : trois de qualité élevée et deux de qualité moyenne);
- c) uniformité des résultats;
- d) caractère direct des données probantes

Tenez compte des résultats de la revue de littérature dans le contexte d'autres documents disponibles.

Les revues systématiques de littérature de haute qualité ont déjà une évaluation critique à profondeur des données probantes tirées d'études pertinentes de sorte que vous n'êtes pas obligé de passer chaque étude en revue vous-même.

OUTIL D'ÉVALUATION CRITIQUE D'UNE REVUE DE LITTÉRATURE

Question clé : _____

Auteur : _____ Année : _____ Numéro de référence : _____

Titre : _____

Évaluateur : _____ Date : _____

Voir le dictionnaire de l'outil d'évaluation critique d'une revue de littérature pour obtenir la liste complète des critères.

À moins d'indication contraire (par l'énoncé « l'un ou l'autre des éléments suivants »), la plupart des critères applicables énumérés pour tous les classements, voire tous ces critères, devraient être présents pour que l'on attribue à l'élément le classement en question.

| Type de revue de littérature | | |
|------------------------------|-----------------------------------|--------------------------|
| Méta-analyse | Revue systématique de littérature | Revue narrative |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

| Question de filtrage | | | |
|--|---|---|--|
| | Forte | Modérée | Faible |
| 1. Questions/cibles de l'analyse claires | Clairement ciblées. Lien solide avec la question clé des lignes directrices. <input type="checkbox"/> | Assez bien ciblées. Lien avec la question clé des lignes directrices. <input type="checkbox"/> | Vagues ou trop générales. Aucun lien avec la question clé des lignes directrices. <input type="checkbox"/> |
| 2. Études incluses et évaluation critique de ces études | Études se rapportant à la question clé incluses. Études analytiques incluses. Critères d'inclusion clairs. Études évaluées de façon uniforme et systématique et résultats clairs. <input type="checkbox"/> | Études pertinentes incluses. Études analytiques incluses. Critères d'inclusion pouvant être vagues. Critères de l'évaluation critique de l'étude vagues, mais résultats de l'évaluation clairs. <input type="checkbox"/> | L'un ou l'autre des éléments suivants : Études pertinentes non incluses. Études analytiques non incluses. Critères d'inclusion sont vagues et absence de résultats de l'évaluation critique pour chaque étude. <input type="checkbox"/> |

Commentaires :

| Décision relative au filtrage | |
|--|--|
| <input type="checkbox"/> Rejeter (si faible) | OU <input type="checkbox"/> Poursuivre |
| *Remarque : Si l'élément no 2 à évaluer est modéré ou fort mais que l'élément no 1 est faible, alors il faut soigneusement déterminer l'intérêt de poursuivre. | |

| Évaluation de la méthode | | | |
|---|---|---|---|
| | Forte | Modérée | Faible |
| 3. Recherche d'études pertinentes | Recherche complet dans plusieurs bases de données, des bibliographies, des articles rédigés dans une langue autre que l'anglais et la littérature grise ou non publiée. <input type="checkbox"/> | Recherche complet dans des bases de données incluant la littérature rédigée dans une langue autre que l'anglais mais susceptible de ne pas avoir tenu compte des bibliographies et de la littérature grise ou non publiée. <input type="checkbox"/> | Recherche limitée dans des bases de données et la littérature rédigée dans une langue autre que l'anglais. On n'a pas tenu compte de la littérature grise ou non publiée. <input type="checkbox"/> |
| 4. Rigueur du processus de revue | Les études incluses respectaient les critères d'inclusion et d'évaluation critique. Elles ont été filtrées et examinées par plus d'un évaluateur, à l'aide des mêmes critères et une bonne entente. <input type="checkbox"/> | Les études incluses respectaient les critères d'inclusion et d'évaluation critique, mais elles ont été filtrées et examinées par un seul évaluateur ou les critères n'étaient pas clairs. <input type="checkbox"/> | Les critères d'inclusion ou d'évaluation critique n'ont pas été utilisés ou l'on ignore si c'était le cas. <input type="checkbox"/> |
| 5. Dans le cas d'une méta-analyse, était-il raisonnable de le faire? Ne s'applique pas <input type="checkbox"/> | Les études combinées ne présentaient pas de différences considérables. Hétérogénéité minimale entre les résultats des études individuelles. Statistiques sommaires appropriées utilisées. <input type="checkbox"/> | Les études combinées ne différaient pas considérablement. Hétérogénéité significative entre les résultats des études, mais les auteurs ont abordé cette question de façon adéquate. Les statistiques semblent raisonnables. <input type="checkbox"/> | Les études combinées différaient considérablement. Hétérogénéité significative entre les résultats des études qui n'a pas été abordée de façon adéquate. Les statistiques ne semblaient pas raisonnables. <input type="checkbox"/> |
| Commentaires : | | | |

| Décision relative à la méthode |
|---|
| <input type="checkbox"/> Rejeter (Si l'élément à évaluer n° 4 est faible, arrêtez l'analyse). Si l'élément n° 3 ou l'élément n° 5 est faible, déterminez soigneusement quel est l'intérêt de poursuivre. Si la revue est interrompue, déterminez les études de la revue de littérature qui sont pertinentes et évaluez-les individuellement. |
| <input type="checkbox"/> Poursuivre . Si les éléments à évaluer n° 3 à 5 sont modérés ou forts, poursuivez l'évaluation. |

| Évaluation des résultats de l'étude (ampleur de l'effet) | | | |
|---|--|--|--|
| | Forte | Modérée | Faible |
| 6. Description et interprétation des résultats de l'étude (allez à la question 7 dans le cas d'une méta-analyse) Ne s'applique pas <input type="checkbox"/> | Interprétation correcte de la signification statistique et de l'intervalle de confiance (IC) ou synthèse raisonnable de la tendance et de l'incidence potentielle. <input type="checkbox"/> | Interprétation correcte de la signification statistique et de l'IC ou synthèse raisonnable de la tendance, mais aucune analyse de l'incidence potentielle. <input type="checkbox"/> | N'a pas correctement interprété les résultats. <input type="checkbox"/> |

| | | | |
|---|--|---|--|
| <p>7. Dans le cas d'une méta-analyse seulement : Ampleur et précision de l'effet du traitement</p> <p>Ne s'applique pas <input type="checkbox"/></p> | <p>IC global de 95 % ou de 99 % signalé. Différence minimale sur le plan de l'ampleur de l'effet du traitement et large chevauchement des IC des études individuelles. Puissance suffisante. Interprétation correcte de la signification statistique et de l'IC.</p> <p style="text-align: center;"><input type="checkbox"/></p> | <p>IC global de 95 % ou de 99 % signalé. Une certaine différence sur le plan de l'ampleur de l'effet du traitement et un certain chevauchement des IC des études individuelles. La puissance semblait suffisante. Interprétation correcte de la signification statistique et de l'IC.</p> <p style="text-align: center;"><input type="checkbox"/></p> | <p>L'un ou l'autre des éléments suivants (même si un IC global de 95 ou 99 % a été signalé) : Grande différence sur le plan de l'ampleur de l'effet du traitement et peu de chevauchement des IC des études individuelles, voire aucun. Puissance insuffisante. Mauvaise interprétation des résultats.</p> <p style="text-align: center;"><input type="checkbox"/></p> |
|---|--|---|--|

| Décision relative aux résultats | |
|--|---|
| Déterminez si les données probantes suffisent à recommander des mesures. | |
| a) Y a-t-il un effet clair? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| b) Les résultats sont-ils uniformes d'une étude à l'autre? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| c) Le nombre d'études ayant contribué à la décision relative à l'effet clair est-il suffisant (au moins quatre)? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| d) Les données probantes sont-elles directes? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| e) L'effet est-il significatif sur le plan clinique? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| f) Dans le cas d'une méta-analyse, les données ont-elles été correctement mises en commun et a-t-on mené une analyse statistique adéquate? | <input type="checkbox"/> Oui <input type="checkbox"/> Non |
| <p>Si la réponse à chaque question est oui, alors il peut être nécessaire d'évaluer l'applicabilité à l'aide des éléments à évaluer nos 8 et 9.</p> <p>Si la réponse à l'une de ces questions est négative, alors n'évaluez pas les éléments 8 et 9; passez à l'élément à évaluer no 10 et tirez une conclusion globale, mais n'énoncez pas de recommandation.</p> | |
| <p>Décision relative au caractère direct des données probantes présentées dans l'étude</p> <p>Tirez une conclusion relative au caractère direct des données probantes :</p> <p><input type="checkbox"/> Les données probantes directes proviennent des études portant précisément sur l'association étudiée.</p> <p><input type="checkbox"/> L'extrapolation est l'inférence tirée d'études qui portaient sur une question différente, mais connexe, ou la même question, mais dans des conditions artificielles.</p> | |
| <p>Commentaires :</p> | |

| Évaluation de l'applicabilité | | | |
|--|--|---|---|
| | Forte | Modérée | Faible |
| 8. Application des résultats à la population étudiée | Population de l'échantillon et contexte très semblables à la population étudiée. <input type="checkbox"/> | Population de l'échantillon et contexte quelque peu semblables à la population étudiée. <input type="checkbox"/> | Population de l'échantillon et contexte différents de la population étudiée. <input type="checkbox"/> |
| 9. Applicabilité à la lumière de résultats importants (p. ex. coûts, perspectives des intervenants) | L'intervention est très susceptible d'être facilement mise en œuvre dans d'autres contextes. <input type="checkbox"/> | L'intervention est quelque peu susceptible d'être facilement mise en œuvre dans d'autres contextes. <input type="checkbox"/> | L'intervention est peu susceptible d'être facilement mise en œuvre dans d'autres contextes. <input type="checkbox"/> |
| <p>Commentaires :</p> <p>Incluez les grandes faiblesses ou limites (p. ex. important manque d'uniformité des résultats, probabilité élevée d'un biais de déclaration, incertitude quant au caractère direct des données probantes).</p> | | | |

| Conclusion globale et tableau de synthèse des données probantes | | | |
|--|-----------------------------|------------------------------|---|
| 10. Peut-on tirer une conclusion à la lumière des données probantes? | | <input type="checkbox"/> Oui | <input type="checkbox"/> Non |
| Si la réponse est non et que vous ne pouvez pas utiliser la revue de littérature dans son ensemble, vérifiez les raisons et évaluez les études individuelles. | Rejet au moment du filtrage | Méthodes de revue faibles | Quantité insuffisante de données probantes pour formuler une recommandation |
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <p>Si la réponse est oui et qu'il y avait assez de données probantes pour formuler une recommandation et que les résultats étaient applicables à la population étudiée, déterminez ce qui suit :</p> <ul style="list-style-type: none"> • Force du plan d'étude (applicable aux méta-analyses seulement) : Forte • Revues systématiques de littérature et revues narratives : Aucun classement <p>Décision relative à la qualité de l'étude</p> <p>La conclusion globale devrait être relative à la qualité (rigueur) des méthodes d'analyse ainsi qu'à la qualité des études de recherche incluses dans la revue de littérature et, par conséquent, à la crédibilité du corpus de données probantes abordées dans la revue de littérature. Avant de formuler une recommandation fondée sur la revue de littérature, il faut déterminer s'il existe une association claire entre l'exposition et le résultat et si les échantillons des études incluses dans la revue de littérature ressemblent au groupe auprès duquel on souhaite généraliser les résultats.</p> <p>Examinez les classements attribués à la méthode et aux décisions relatives aux résultats.</p> <p><input type="checkbox"/> Classez la qualité dans la catégorie ÉLEVÉE si la décision relative à la méthode était forte et que la conclusion globale par rapport à l'association entre l'exposition et le résultat étudié provient de quatre études, ou plus, ayant une force du plan d'étude forte et dont la qualité est élevée.</p> <p><input type="checkbox"/> Classez la qualité dans la catégorie MOYENNE si les méthodes d'analyse étaient classées dans la catégorie modérée ou que les méthodes ont été classées dans la catégorie forte, mais que moins de quatre études ont contribué à la conclusion globale ou que la force des plans d'étude incluses n'était pas solide et que leur qualité n'était pas élevée.</p> <p>Toute revue de littérature ayant une méthode faible devrait être considérée comme de faible qualité et aurait dû être rejetée du processus d'évaluation.</p> | | | |
| Formulez une recommandation : | | | |
| Commentaires : | | | |
| Remarque : Selon la politique des trois Conseils, l'évaluation de programmes et les études de surveillance ne doivent pas faire l'objet d'une approbation éthique. | | | |

| Remplir le tableau des données probantes |
|---|
| <p>Lorsque les données probantes étaient suffisantes à la formulation d'une recommandation et que les résultats étaient applicables à la population étudiée :</p> <p>Synthétisez ce qui suit et ajoutez le résultat au tableau de synthèse des données probantes (à l'aide des définitions pour évaluer les données probantes) :</p> <ol style="list-style-type: none"> conclusion globale et résultats relatifs à l'effet; nombre d'études pour chaque catégorie de plan et qualité de la revue systématique de littérature (p. ex. cinq études ayant une force du plan d'étude forte : trois de qualité élevée et deux de qualité moyenne); uniformité des résultats; caractère direct des données probantes <p>Tenez compte des résultats de la revue de littérature dans le contexte d'autres documents disponibles.</p> |
| Commentaires : |

Les revues systématiques de littérature de haute qualité ont déjà une évaluation critique des données probantes tirées d'études pertinentes de sorte que vous n'êtes pas obligé de passer chaque étude en revue vous-même.

ANNEXE A : GLOSSAIRE, LISTE DES ABRÉVIATIONS ET SYNTHÈSE DES TESTS STATISTIQUES COMMUNS

Les définitions qui suivent ont été tirées ou adaptées des ouvrages de référence suivants :

1. Field A. *Discovering statistics using SPSS*, 2^e Éd. SAGE Publications Ltd., Londres, 2005.
2. Hennekens, CH et JE Buring. *Epidemiology in Medicine*, dir. S. Mayrent, Little Brown and Company, Boston (É.-U.), 1987.
3. Hopkins WG. *A New View on Statistics*, 2009. <http://www.sportsci.org/resource/stats/contents.html>
4. Association internationale d'épidémiologie. *Dictionary of Epidemiology*, 3^e éd., dir. JM Last, Presses de l'Université d'Oxford, 1995.
5. Keyton, J. *Communication Research: Asking Questions, Finding Answers*, 2^e éd. University of Kansas Online Learning Centre. http://highered.mcgraw-hill.com/sites/0073049506/student_view0/glossary.html
6. Nordness, Robert. *Epidemiology and Biostatistics Secrets*. Elsevier, Philadelphie, 2006.
7. StatSoft Inc. © *Glossary, Electronic Statistics Textbook*. Tulsa (OK, É.-U.). <http://www.statsoft.com/textbook/statistics-glossary/t/button/t/>
8. SUPPORT. *Supporting Policy Relevant Reviews and Trials*, « Glossary ». <http://www.support-collaboration.org/summaries/explanations.htm>

Remarque : Pour obtenir de plus amples renseignements, consultez les documents de référence énumérés ci-dessus ou dans la bibliographie.

Ampleur (ou taille) de l'effet¹

L'effet est un terme générique qui renvoie au « résultat d'une cause ». Lorsque nous mesurons l'ampleur d'un effet (par une manipulation expérimentale ou la force d'une relation entre variables), il est question de l'ampleur de l'effet. L'ampleur de l'effet est une mesure de l'ampleur (ou du degré) d'un effet observé.

Analyse à plusieurs variables⁶

Évalue la relation entre un résultat et plusieurs facteurs ou variables prédictives. La régression (p. ex. multiple ou logistique), l'analyse de survie et l'analyse de la variante sont les types les plus courants d'analyse à plusieurs variables.

Analyse bivariée⁶

Évalue la relation entre un résultat et une variable prédictive. Les tableaux croisés et le calcul du rapport de cotes ou du risque relatif à partir d'un tableau 2X2 sont les types d'analyse bivariée les plus courants.

| | |
|--|--|
| Analyse de sensibilité⁴ | Analyse utilisée pour déterminer la mesure dans laquelle les résultats d'un essai ou d'une revue systématique de littérature sont sensibles aux changements liés à la méthode du procédé. Elle examine à quel point les résultats sont touchés par les changements liés aux méthodes, aux valeurs des variables ou aux hypothèses. |
| Analyse multivariée | Évalue la relation entre de multiples résultats et de multiples variables prédictives. L'analyse discriminante et l'analyse de la variance à plusieurs variables sont des exemples d'analyse multivariée. |
| Analyse univariée⁶ | Décrit la fréquence d'un facteur ou d'une variable unique, par exemple, l'âge, le sexe ou les habitudes de tabagisme. La variable peut avoir un certain nombre de catégories (p. ex. n'a jamais fumé, arrêt récent, fume actuellement < un paquet par jour ou fume actuellement > un paquet par jour) ou une variable continue (p. ex. âge). |
| Appariement⁶ | La sélection de cas et de témoins (ou de personnes qui feront partie du groupe d'intervention ou du groupe témoin) qui possèdent des caractéristiques semblables mises en correspondance, comme l'âge, le sexe ou l'emploi. Dans le cas d'appariement individuelle (un pour un), chaque cas ou sujet du groupe d'intervention a son équivalent du groupe témoin qui possède les mêmes caractéristiques. Dans le cas d'appariement de groupes, la proportion de témoins possédant une caractéristique donnée est identique à la proportion de cas possédant la même caractéristique, mais il n'y a pas eu de sélection individuelle. |
| Coefficient de corrélation de Pearson⁴ | Mesure d'association qui indique le degré de la relation linéaire (corrélation) entre deux variables continues. Il s'agit du coefficient de corrélation le plus couramment utilisé. Le coefficient de corrélation peut aller de +1 (témoignant d'une relation positive parfaite), jusqu'à -1 (témoignant d'une relation négative parfaite) en passant par zéro (témoignant de l'absence d'une relation). En général, un coefficient de corrélation qui varie entre 0,00 et 0,3 est considéré comme étant faible, lorsqu'il varie entre 0,3 et 0,7, il est considéré comme étant modéré et, lorsqu'il varie entre 0,7 et 1, il est considéré comme étant élevé. Toutefois, cette règle devrait toujours être interprétée selon les circonstances. Il faut noter que la corrélation ne témoigne pas d'une relation de cause à effet. |

| | |
|------------------------------|---|
| Confusion⁴ | Distorsion de la relation entre une exposition et un résultat en fonction des effets de différents facteurs. Dans un contexte épidémiologique, les variables confusionnelles entraînent une confusion des effets de l'exposition sur le résultat et des effets des facteurs sur le résultat. Pour qu'une variable soit confusionnelle, le facteur doit être associé à l'exposition et au résultat et être exclu de la chaîne causale. |
| Données³ | Les données désignent des groupes de renseignements qui représentent les caractéristiques qualitatives ou quantitatives d'une variable. Habituellement, il s'agit d'un ensemble de chiffres, de caractères, d'images ou d'autres extrants prélevés à l'aide de mécanismes qui mesurent ou recueillent des données. Les variables qui correspondent à un chiffre sont appelées des variables numériques ou des données numériques ; celles désignées par un nom ou une étiquette, sans être assorties d'un ordre ou d'un rang, sont des variables nominales ou des données nominales . Les variables qui portent un nom ou une étiquette et qui s'inscrivent manifestement dans un ordre ou une hiérarchie sont des variables ordinales ou des données ordinales . Les variables nominales et ordinales regroupées en catégorie sont aussi appelées des données catégoriques . Une variable numérique qui comporte des intervalles équivalents est appelée une donnée d'intervalle en l'absence d'un point zéro significatif et appelée une donnée de rapport s'il y a un point zéro significatif. |
| Données agrégées | Données déclarées pour un groupe ou une population à titre d'unité plutôt que pour des membres du groupe ou de la population (données individuelles). |
| Données catégoriques | Voir Données. |
| Données de rapport | Voir Données. |
| Données d'intervalle | Voir Données. |
| Données nominales | Voir Données. |
| Données numériques | Voir Données. |
| Données ordinales | Voir Données. |

| | |
|--|--|
| Écart-type⁸ | Mesure de la variation ou de la dispersion d'un ensemble d'observations, qui correspond à la racine carrée positive de la variance. |
| Échantillon⁴ | Segment sélectionné d'une population donnée (p. ex. tous les patients traités pour brûlures qui développent un quelconque type d'infection). Un échantillon peut être aléatoire ou non aléatoire (c.-à-d. sélectionné à l'aide d'une méthode aléatoire ou non aléatoire) et peut être représentatif ou non représentatif (c.-à-d. posséder ou ne pas posséder de caractéristiques semblables à la population cible). |
| Échantillonnage⁴ | Processus de sélection d'un certain nombre de sujets à partir de tous les sujets d'un groupe particulier. |
| Échantillonnage aléatoire^{4,5,8} | Sélection des participants à une étude par l'intermédiaire d'un processus aléatoire, de sorte que la probabilité d'être sélectionné soit équivalente pour tous les participants potentiels. Ce processus n'a pas d'incidence directe sur la confusion, mais il contribue à la possibilité de généraliser les résultats. |
| Étude de cohorte^{4,8} | Étude d'observation qui porte sur le suivi d'un groupe défini de personnes sur une certaine période. Les résultats sont comparés entre les sujets exposés et non exposés (ou entre des sujets exposés à différents degrés) à une intervention particulière ou à un autre facteur étudié. Une étude de cohorte prospective rassemble des participants, puis les suit. Une étude de cohorte rétrospective définit des sujets à partir de dossiers et en assure le suivi à partir d'un moment précisé au préalable jusqu'au présent ou à la fin d'une période de collecte de données précisée au préalable. |
| Étude d'intervention² | Étude supposant la comparaison des résultats entre au moins deux groupes délibérément soumis à une intervention (habituellement un traitement, mais parfois une mesure préventive, comme la vaccination) pour vérifier une hypothèse. |
| Étude par observation² | Type d'étude dans le cadre de laquelle les sujets sont observés ou certains résultats sont mesurés. L'intervention ou le facteur de risque est survenu naturellement, et il n'y a aucune tentative d'influencer le résultat. |

| | |
|--|--|
| Étude sur le lien épidémiologique | Catégorie d'études qui regroupe les enquêtes sur les dons antérieurs, le retraçage et les contacts. Dans le cadre de cette étude, les sujets font l'objet d'une évaluation pour trouver un lien (p. ex. contact ou typage microbien) avec des cas, des contacts ou des conditions. |
| Exposition | Le terme exposition est utilisé largement dans cette trousse d'outils pour référer aux expositions étudiées tel que les facteurs de risque, les facteurs de protection ou démographiques, et les interventions. |
| Fiabilité⁴ | Mesure dans laquelle il est possible de reproduire les résultats obtenus par l'intermédiaire d'une mesure ou d'une procédure, qui témoigne du degré d'uniformité ou de répétabilité. Le manque de fiabilité peut être le résultat d'un instrument de mesure, de la variation entre les observateurs ou de l'instabilité de la caractéristique mesurée. |
| Fidélité interévaluateurs⁸ | Variation des mesures prises par différentes personnes à l'aide de la même méthode ou du même instrument. |
| Graphique en forêt (Forest plot) | Représentation graphique des résultats individuels de chaque étude d'une méta-analyse et du résultat combiné de la méta-analyse. Le schéma permet au destinataire de voir l'hétérogénéité des résultats des études. Les résultats d'études individuelles sont désignés par des carrés alignés sur l'estimation ponctuelle de chaque étude. Une ligne horizontale traverse tous les carrés pour montrer l'intervalle de confiance de chaque étude – habituellement 95 %, mais pas toujours. L'estimation globale provenant de la méta-analyse et l'intervalle de confiance sont représentés par un diamant. Le centre du diamant correspond aux estimations ponctuelles mises en commun et les extrémités horizontales correspondent à l'intervalle de confiance. |
| Hétérogénéité⁸ | Variation ou diversité des participants, des interventions et des mesures des résultats dans le cadre d'une étude ou d'un ensemble d'études. Un ensemble d'études ou de participants dont l'hétérogénéité est mesurable est dit hétérogène (le contraire d'homogène). |
| Homogénéité⁸ | Employé dans son sens général, le terme signifie que les participants, les interventions ou les mesures des résultats sont semblables à l'échelle d'un ensemble d'études ou dans le cadre d'une même étude. |

| | |
|--|--|
| Hypothèse nulle ^{3,4} | En termes simples, selon l'hypothèse nulle, les résultats observés dans une étude, une expérience ou un essai sont identiques à ce qui se serait produit purement par hasard. |
| Intervalle ⁴ | Ensemble qui contient tous les chiffres entre deux chiffres donnés. |
| Intervalle de confiance (IC) ⁴ | Intervalle défini à partir d'une probabilité donnée (p. ex. 95 % ou 99 %) que la véritable valeur d'une variable, comme la moyenne, la proportion ou le taux, soit contenue dans l'intervalle. Lorsqu'un intervalle de confiance est étroit, l'estimation de la valeur du paramètre est réputée plus précise. Plus la variation de l'échantillon est grande, plus l'IC sera large. Les limites de l'IC sont nommées limites de confiance. La largeur de l'IC désigne la différence entre les limites de confiance. |
| Littérature grise | La littérature qui n'est pas publiée par des maisons d'édition ou indexée dans des bases de données d'articles de revues telles que <i>PubMed</i> ou <i>CINAHL</i> . Les documents gouvernementaux et les comptes rendus des conférences sont des sources communes de littérature grise. |
| Niveau alpha ⁵ | Le niveau alpha représente la probabilité d'une erreur de type I, ce qui correspond au rejet de l'hypothèse nulle alors qu'elle est vraie (et aurait donc dû être acceptée). Lorsqu'une erreur de type I se glisse, on conclut qu'il existe une différence alors qu'il n'y en a pas. Dans le cadre d'un test d'hypothèse, alpha sert de seuil pour décider si une valeur p a une signification statistique ou non. La valeur attribuée à alpha est quelque peu arbitraire, mais, habituellement, elle est établie à 0,05, parfois, à 0,01. |
| Paramètre ⁴ | Dans le domaine de l'épidémiologie et de la statistique, il s'agit d'une caractéristique mesurable chez une population qui est souvent estimée à l'aide d'une statistique, p. ex. moyenne, écart-type, rapport de cotes, etc. |
| Participants | Voir Participant à l'étude. |
| Participants à l'étude ⁶ | Personnes ou échantillons qui font l'objet d'une étude. Les participants sont souvent des patients, mais pas nécessairement, et sont généralement sélectionnés dans une population d'intérêt particulière. Voir aussi Échantillon. |
| Plan d'étude ⁴ | « Architecture » d'une étude : sa structure, les détails particuliers à la population étudiée, l'échéancier, les méthodes, les procédures et les considérations éthiques. |

| | |
|---|---|
| Population (dans l'échantillonnage) ^{4,8} | Toutes les unités ou personnes auprès desquelles on peut prélever un échantillon. Une population peut être définie par une quelconque caractéristique, p. ex. géographie, tranche d'âge, certaines maladies, institutions, dossiers ou événements. L'échantillon a pour but de donner des résultats représentatifs de toute la population. |
| Population cible ⁴ | Groupe à partir duquel une population étudiée est sélectionnée ou population chez laquelle on a l'intention d'appliquer les résultats de l'étude. Les inférences et les recommandations découlant de l'étude risquent d'être moins valides si elles sont appliquées à une population qui comporte différentes caractéristiques (p. ex. âge, problèmes de santé, situation sociale, etc.) par rapport à la population étudiée. |
| Puissance ⁴ | Probabilité de la démonstration d'une association statistiquement significative si elle existe. La puissance d'une étude est déterminée par l'ampleur de l'effet, la dispersion de la population ou de l'échantillon et la taille de l'échantillon. |
| Question clé ⁴ | Question axée sur l'enjeu fondamental à aborder dans le cadre de l'évaluation critique. Au moment de l'élaboration de lignes directrices, on a besoin d'une série de questions clés structurées pour bien définir le contenu des lignes directrices en fonction de la portée et des objectifs définis. |
| Rapport de cotes (RC) ⁷ | Cotes d'un événement dans un groupe par rapport aux cotes de cet événement dans un autre groupe. Il y a plusieurs types de rapport de cotes et des formules différentes. Le RC généralement le plus utilisé, surtout à l'intérieur des études cas/témoins, est obtenu par la division des cotes d'exposition du groupe avec le résultat (cas) par les cotes d'exposition du groupe sans le résultat (témoins). Un rapport de cotes de 1 révèle qu'il n'y a pas de différence entre les groupes comparés. Un RC inférieur à 1 révèle que l'association entre l'exposition et le résultat est inférieure chez les cas comparée aux témoins, tandis qu'un RC supérieur à 1 révèle que l'association entre l'exposition et le résultat est supérieure chez les cas comparée aux témoins. Le RC est une estimation du risque relatif (RR) mais peut entraîner une surestimation du risque. Toutefois, lorsque le risque est faible, le RC sera très semblable au RR. |
| Ratio ⁴ | Un chiffre relativement à un autre. Le taux, la proportion et le pourcentage sont des exemples des rapports les plus couramment utilisés. |

| | |
|--|---|
| Régression linéaire³ | Forme d'analyse de la régression qui définit un résultat comme étant la fonction d'au moins un facteur. La variable du résultat est une variable continue. Une régression linéaire simple établit le modèle d'une seule variable prédictive ou d'un seul facteur de risque, tandis qu'une régression linéaire multiple établit le modèle de plusieurs variables prédictives ou facteurs de risque. |
| Régression logistique⁸ | Forme d'analyse de régression qui établit le modèle d'une probabilité individuelle d'avoir une maladie ou de présenter un autre résultat à la suite d'au moins un facteur de risque ou d'au moins une variable prédictive. La variable du résultat est dichotomique, c.-à-d. qu'elle peut aboutir à seulement deux résultats, comme mort ou vivant. |
| Répartition aléatoire (randomisation)^{3,8} | Processus de répartition aléatoire des participants dans l'un des groupes d'un essai contrôlé. Les deux composantes de la répartition aléatoire sont la génération d'une séquence aléatoire et sa mise en œuvre. La probabilité d'être affecté à un groupe devrait être la même que la probabilité d'être affecté à l'autre groupe. Idéalement, la mise en œuvre est faite de sorte que les participants admis à l'étude ignorent la séquence (dissimulation de l'affectation des sujets aux deux groupes). L'objectif est de répartir également les variables confusionnelles inconnues entre les groupes. |
| Répartition quasi aléatoire⁸ | Dans le cadre d'un essai, méthodes de répartition des personnes dans des groupes qui ne sont pas aléatoires, mais conçues pour produire des groupes semblables. Voici des exemples de méthodes quasi aléatoires : répartition selon la date de naissance de la personne, selon le jour de la semaine ou le mois de l'année, selon le numéro de dossier médical de la personne ou en répartissant simplement une personne sur deux. |
| Résultat | Le terme résultat est utilisé largement dans cette trousse d'outils pour référer aux résultats étudiés tels que les infections, maladies, comportements, effets ou affections. |
| Risque relatif ou rapport de risque⁸ | Risque d'un groupe par rapport à un autre. Dans le cadre d'une étude d'intervention, il s'agit du risque dans le groupe d'intervention/exposé par rapport au risque du groupe témoin/non-exposé. Un risque relatif (RR) de 1 témoigne de l'absence de différence entre les groupes comparés. Un RR inférieur à 1 révèle que l'exposition a réduit le risque de ce résultat, tandis qu'un RR supérieur à 1 révèle que l'exposition a augmenté le risque de ce résultat. |

| | |
|--|--|
| Schéma d'expériences croisées⁴ | Type d'étude d'intervention qui compare au moins deux interventions (ou une intervention et aucune intervention) dans le cadre de laquelle les participants, après avoir subi une intervention, font l'objet de l'autre intervention. On reproche à ce plan d'entraîner le risque que les effets du premier traitement soient transférés à la période du deuxième traitement. |
| Signification statistique^{3,4} | L'approche traditionnelle à la déclaration d'un résultat exige que l'on détermine s'il est statistiquement significatif. Pour ce faire, on génère une valeur p à partir d'une variable à tester . La valeur p obtenue est alors comparée au niveau alpha présélectionné (p. ex. 0,05 ou 0,01). Si la valeur p obtenue est inférieure à alpha (p. ex. $p < 0,05$), le résultat est réputé statistiquement significatif, ce qui indique qu'il est peu probable ($< 5\%$) que le résultat soit le produit du hasard uniquement. Les résultats sont statistiquement significatifs ou non; une valeur p de 0,01 n'est pas « plus significative » qu'une valeur p de 0,04. |
| Tendance centrale⁵ | Les conclusions liées à la signification statistique peuvent aussi être déterminées à l'aide de l'intervalle de confiance. Si l'IC d'un RC ou d'un RR n'inclut pas le chiffre 1 ou que l'IC correspondant à la différence des moyennes n'inclut pas le chiffre 0, l'effet est réputé statistiquement significatif. Il faut noter que la signification statistique n'entraîne pas automatiquement l'importance clinique. |
| Test t⁷ | Terme appliqué à l'une ou l'autre des mesures qui résument la distribution de notes (ou l'ensemble de valeurs). La moyenne, la médiane et le mode sont des mesures communes de la tendance centrale; ce chiffre unique fait office de synthèse de toutes les valeurs d'une variable. |
| Valeur p⁴ | Le test t est la méthode la plus couramment utilisée pour évaluer les différences de moyenne entre deux groupes. Le test t non apparié est employé lorsque les mesures sont indépendantes (p. ex. pression sanguine de patients à qui l'on a administré un médicament comparativement à un groupe témoin à qui l'on a administré un placebo), tandis qu'un test t apparié est employé si les mesures sont dépendantes (p. ex. pression sanguine de patients « avant » comparativement à « après » l'administration d'un médicament). Théoriquement, on peut employer le test t même si les échantillons sont très petits (p. ex. seulement dix sujets), à condition que les variables soient distribuées approximativement normalement et que la variation de la cote des deux groupes ne comporte pas de différences sur le plan de la fiabilité. |

Validité^{4,8}

Probabilité qu'une variable à tester soit aussi extrême ou plus extrême que les observations si l'hypothèse nulle était vraie. Il s'agit d'un énoncé de la probabilité que la différence observée soit le produit du hasard si les groupes étaient réellement semblables (selon l'hypothèse nulle).

Variable dichotomique²

La validité d'un instrument est l'expression de la mesure dans laquelle l'instrument ou la procédure mesure ce qu'il a été conçu pour mesurer. La validité interne d'une étude est la mesure dans laquelle les résultats de l'étude sont susceptibles d'être vrais et exempts de biais (erreurs systématiques). La validité externe d'une étude est la mesure sur laquelle les résultats de l'étude peuvent être généralisés à d'autres populations ou contextes.

En statistique, une variable dichotomique désigne une variable associée à seulement deux possibilités (p. ex. mort ou vivant).

| Liste des abréviations | |
|------------------------|-----------------------------------|
| ECAA | Étude comparative avant-après |
| ECNR | Essai contrôlé non randomisé |
| ECR | Essai contrôlé randomisé |
| ENAA | Étude non comparative avant-après |
| IC | Intervalle de confiance |
| OEC | Outil d'évaluation critique |
| RC | Rapport de cotes |
| RR | Risque relatif |
| STI | Séries temporelles interrompues |

Tableau 5 – Synthèse des tests statistiques communs

Afin de déterminer si un test statistique approprié a été employé, il faut tout d'abord définir le type de données incluses et le nombre de variables d'exposition (prédictives) étudiées. Voici la synthèse des tests les plus couramment utilisés.

| Test | Nombre de variables | Type de données | Description des statistiques |
|--|---------------------|--------------------------------|--|
| Statistiques descriptives | | | |
| Moyenne et écart type (ET) | 1 | Intervalle ou ratio | Sert à l'analyse des tendances centrales et de la dispersion. L'écart-type décrit la variation ou la dispersion des valeurs par rapport à la moyenne; 95 % des valeurs $\pm 2,6$ ET et 99,7 % $\pm 3,0$ ET. |
| Risque relatif (RR) | 2 | Dichotomiques | Sert à mesurer le risque du résultat dans le groupe exposé relativement au groupe non exposé. Il n'est utilisé que dans les études de cohorte ou les études interventionnelles. |
| Rapport de cotes (RC) | 2 | Dichotomiques | Sert à mesurer les cotes d'exposition dans les cas relatifs aux témoins. Sert aussi à estimer le RR; peut être utilisées dans n'importe quel type d'étude. |
| Médianes et écart interquartile (EIQ) | 1 | Ordinales, intervalle ou ratio | Sert à analyser les tendances centrales. La médiane est employée si l'intervalle ou le ratio est asymétrique. La médiane est la valeur qui divise le groupe en deux parties égales : 50 % des valeurs s'inscrivent sous la médiane. L'EIQ représente 50 % des valeurs du centre. |
| Statistiques corrélationnelles | | | |
| Statistique kappa de Cohen ou Kappa pondéré | >1 | Nominales ou ordinales | Sert à évaluer la fidélité interévaluateurs. Le kappa pondéré est employé dans le cas où il y a de multiples utilisateurs et de multiples résultats. |
| Coefficient de corrélation de Pearson | >1 | Intervalle ou ratio | Sert à mesurer la corrélation entre deux variables continues. |
| Statistiques inférentielles | | | |
| Khi carré (χ^2) ou test exact de Fisher | >1 | Catégoriques | Sert à mesurer l'écart entre la distribution de la fréquence observée et attendue. Le khi carré est une estimation du test exact de Fisher. Les deux procédés évaluent les différences sur le plan de la fréquence, des proportions, du rapport de cotes ou du risque relatif. Le khi carré de McNemar est employé pour les données appariées. |

| Test | Nombre de variables | Type de données | Description des statistiques |
|--|--------------------------|--|--|
| Régression logistique en l'absence d'un facteur temporel Risques proportionnels de Cox, dans le cas d'un facteur temporel | >1 | Résultats : dichotomiques Prédictives : catégoriques ou continues | La régression logistique sert à prévoir la probabilité de la fréquence d'un événement en inscrivant les données sur une courbe logistique. Elle estime le rapport de cotes pour chaque variable prédictive étudiée tout en contrôlant les effets des autres variables prédictives dans le modèle. La régression logistique conditionnelle est employée pour les données appariées. Le modèle de Cox donne un rapport de risque qui est une estimation du RR d'une unité particulière dans le temps. Le modèle de Cox peut aussi être utilisé pour comparer les moments et les événements. |
| Test t ou test t apparié Analyse de la variance (ANOVA) | 2 (t-test) ≥2 (ANOVA) | Résultats : continues Prédictives : catégoriques | Sert à mesurer les différences de moyenne entre deux groupes. Le test t est employé si les mesures sont indépendantes, tandis que le test t apparié est employé si les mesures sont dépendantes. L'analyse de la variance sert à comparer plus de deux moyennes. |
| Régression linéaire multiple (RLM) Analyse de covariance (ANCOVA) | >1 | Résultats : continues Prédictives : catégoriques ou continues | Sert à prévoir un résultat produit par au moins deux facteurs. L'analyse de covariance est une variation de la RLM dans le cadre de laquelle les variables catégoriques sont traitées de façon à permettre l'évaluation de catégories individuelles. |

Remarque : Les variables prédictives peuvent désigner une exposition étudiée. Voir le glossaire pour obtenir les définitions. Consulter un ouvrage de référence en matière de statistique ou d'épidémiologie (mentionné dans le glossaire ou dans la bibliographie) au besoin.

ANNEXE B : EXEMPLE D'UN TABLEAU DE SYNTHÈSE DES DONNÉES PROBANTES ET RECOMMANDATIONS

Tableau 6 – Exemple d'un tableau de synthèse des données probantes et de recommandations

Question clé : Les désinfectants pour les mains à base d'alcool (DMBA) sont-ils une mesure d'hygiène des mains efficace dans un environnement de soins de santé?

| Auteur (année) Numéro de référence | Méthodes et mesures des résultats pertinentes | Résultats | Conclusions Commentaires de l'évaluateur Classement de l'étude |
|---|--|---|--|
| Picheansathian (2004) n° 13369 | Revue systématique de littérature bien effectué (qualité élevée) | Mention de multiples autres études non incluses ici, témoignant de résultats uniformes quant à la réduction de la charge microbienne grâce aux DMBA (différentes concentrations) comparativement à d'autres solutions et quant à l'observation accrue des mesures d'hygiène des mains. | Multiples études dont la force du plan d'étude est forte et la qualité est élevée |
| Larson (2001) N° 8144 | 1 ^{er} groupe : 2 % de lavage au CHG 2 ^e groupe : DMBA (61 % d'éthanol) Mesure de la condition de la peau et de la microbiologique cutanée. 2 USI 50 volontaires (différents types de TS) Dix jours ouvrables, HM et contact avec les patients documentés, techniques d'HM et carnets validés. Cultures de référence, jour 1, fin des semaines 2 et 4. | Aucune différence significative dans la réduction de la quantité de bactéries entre les deux groupes, mais le nombre de bactéries a beaucoup diminué par rapport aux données de référence dans les deux groupes. Procédure avec DMBA beaucoup moins longue qu'avec CHG. Améliorations des conditions cutanées dans le groupe des DMBA. Réduction de 50 % des coûts pour le groupe de DMBA. | ECR Force du plan d'étude forte Qualité élevée On conclut que les DMBA ne sont pas mieux que le LM avec un savon antiseptique pour réduire le nombre d'UFC, mais comptent d'autres avantages. Il n'y a pas eu de comparaison entre les DMBA et le LM avec savon ordinaire. |

| Auteur (année) Numéro de référence | Méthodes et mesures des résultats pertinentes | Résultats | Conclusions Commentaires de l'évaluateur Classement de l'étude |
|---|--|---|---|
| Zaragoza (1999) N° 12753 | <p>Quatre unités de soins et trois USI</p> <p>Échantillon aléatoire de 43 TS prélevés auprès de 175</p> <p>Attribution aléatoire au LM ordinaire ou aux DMBA</p> <p>Chaque groupe a suivi les deux procédures.</p> <p>Cultures prélevées à trois différents moments, avant et après l'HM.</p> <p>DMBA = Sterillium</p> | <p>Réduction significative du nombre d'UFC dans le groupe des DMBA (de 88,2 %) relativement au groupe de LM ordinaire (de 49,6 %), $p < 0,001$.</p> <p>Aucune différence significative entre les groupes sur le plan du nombre d'UFC 30 minutes après l'HM (les DMBA n'ont pas d'effet durable).</p> <p>L'acceptation des DMBA était « bonne », correspondant à 72 % des TS.</p> | <p>ECR prospective avec schéma d'expériences croisées</p> <p>Force du plan d'étude forte</p> <p>Qualité élevée</p> |
| Larson (1986) N° 784 | <p>50 volontaires (pas dans le contexte des soins de santé) ont été attribués de façon aléatoire aux cinq groupes suivants :</p> <ul style="list-style-type: none"> • LM avec savon ordinaire, et quatre groupes témoins – deux DMBA différents (isopropyle, 60 % ou 70 %), un alcool et un antiseptique. • Lavage 15 fois par jour pendant cinq jours et prélèvement de cultures. | <p>DMBA comparativement au savon : à la fin du jour 1 (15 épisodes d'HM), >2 réduction de la quantité de bactéries chez les utilisateurs de DMBA par rapport au groupe de LM avec savon ordinaire.</p> <p>Les utilisateurs préféraient le CHG (moins irritant pour la peau).</p> | <p>Étude en laboratoire</p> <p>Force du plan d'étude forte</p> <p>Qualité élevée</p> <p>Généralisation à un contexte clinique pas facile</p> |
| Bischoff (2000) N° 12562 | <p>Contrôle de base de l'HM, puis programme d'éducation/rétroaction dans deux USI et programme de pression sociale dans l'unité de soins générale, puis introduction de DMBA.</p> <p>Observation directe de l'HM</p> <p>DMBA = 60 % d'alcool, type non précisé</p> | <p>1 575 observations sur 120 jours</p> <p>1 distributeur de DMBA pour 4 lits : l'HM était de 19 % avant le contact avec le patient et de 41 % après le contact avec le patient.</p> <p>1 distributeur de DMBA pour 1 lit : l'HM était de 23 % et de 48 %, respectivement.</p> <p>HM de base : <16 % et < 25 % respectivement.</p> <p>Les différences étaient statistiquement significatives.</p> | <p>Étude non comparative avant-après</p> <p>Force du plan d'étude faible</p> <p>Qualité élevée</p> <p>Selon la conclusion, l'accessibilité des DMBA a contribué à l'HM davantage que l'éducation.</p> |

| Auteur (année) Numéro de référence | Méthodes et mesures des résultats pertinentes | Résultats | Conclusions Commentaires de l'évaluateur Classement de l'étude |
|---|---|---|---|
| Pittet (2000) N° 6630 | Promotion de l'éducation en matière d'HM, rotation d'affiches, points de service pour DMBA/DMBA personnels, rétroaction relative au rendement Participation des responsables de l'administration Vérifications de l'HM répétés au fil du temps DMBA : 75 % d'isopropyle avec CHG | Interventions multiples sur quatre ans à la source d'une observation accrue de la procédure d'HM (pour le personnel infirmier, pas médical) : 48 % (données de base), et 66 % en 1999 ($p < 0,001$) On a aussi conclu que le taux d'ILSS avait beaucoup diminué. | Étude non comparative avant-après Force du plan d'étude faible Qualité élevée |

Remarque : Consulter le tableau 1 (*Définitions des termes utilisés pour évaluer les données probantes*) et le tableau 4 (*Consignes de l'Agence pour classer les données probantes à la source des recommandations*) pour obtenir de plus amples renseignements au sujet de l'attribution d'un classement aux données probantes et de l'évaluation des recommandations.

Texte de synthèse pour la question clé

Recommandation : Les DMBA sont la mesure d'hygiène des mains préférée dans tous les environnements de soins de santé.

Classement des données probantes : **AI**

Justification du classement des données probantes : De multiples études ayant une force de plan d'étude forte et étant de qualité élevée, résultats uniformes se rapportant tous directement à l'efficacité des DMBA en ce qui concerne la réduction de la quantité de bactéries sur les mains dans un environnement clinique, avec l'appui d'études additionnelles dont le plan et la qualité sont inférieurs, mais les résultats, uniformes. Les études confirment aussi que les DMBA augmentent l'observation de l'HM.

Aucune question d'analyse concernant la faisabilité de la mise en œuvre de la recommandation.

ANNEXE C : MEMBRES DU GROUPE DE TRAVAIL D'EXPERTS EN PRÉVENTION ET EN CONTRÔLE DES INFECTIONS

Membres participant au Groupe de travail d'experts en prévention et en contrôle des infections lors de l'élaboration du présent document (anciennement appelé Comité directeur chargé de l'élaboration des lignes directrices pour la prévention et le contrôle des infections) :

- **D^{re} Donna Moralejo**, professeure agrégée, Memorial University School of Nursing, St. John's, Terre-Neuve-et-Labrador (chef de projet)
- **D^{re} Lynn Johnston**, professeure de médecine, QEII Health Science Centre, Halifax, Nouvelle-Écosse (présidente)
- **M^{me} Sandra Boivin**, BSc, agente de planification, de programmation et de recherche, Direction de la santé publique des Laurentides, Saint-Jérôme, Québec
- **Nan Cleator**, IA, consultante nationale en matière d'exercice de la profession, VON Canada, Huntsville, Ontario
- **Brenda Dyck**, BSclnf, CIC, directrice de programme, Programme de prévention et de lutte contre les infections, Office régional de la santé de Winnipeg, Winnipeg, Manitoba
- **D^r John Embil**, directeur, Unité de contrôle des infections, Centre des sciences de la santé, Winnipeg, Manitoba
- **Karin Fluet**, IA, BSclnf, CIC, directrice exécutive, IPC Edmonton Zone and Standards and Projects, Alberta Health Services, Edmonton, Alberta
- **D^{re} Bonnie Henry**, médecin épidémiologiste et professeure adjointe, School of Population & Public Health, Université de la Colombie-Britannique, BC Centre for Disease Control, Vancouver, Colombie-Britannique
- **Dany Larivée**, BSc, coordonnateur de la prévention des infections, Hôpital Montfort, Ottawa, Ontario
- **Mary LeBlanc**, IA, BSclnf, CIC, Tyne Valley, Île-du-Prince-Édouard
- **D^{re} Anne Matlow**, directrice de la prévention des infections, Hospital for Sick Children, Toronto, Ontario
- **D^{re} Dorothy Moore**, Division des maladies infectieuses, Hôpital de Montréal pour enfants, Montréal, Québec
- **Filomena Pietrangelo**, BSclnf, directrice, Secteur de la prévention, Santé et sécurité au travail, Centre universitaire de santé McGill, Montréal, Québec
- **JoAnne Seglie**, IA, COHN-S, chef, Santé au travail, Campus Université de l'Alberta, Office of Environment Health/Safety, Edmonton, Alberta
- **D^r Pierre St-Antoine**, Centre des sciences de la santé, Centre hospitalier de l'Université de Montréal, Hôpital Notre-Dame, Microbiologie, Montréal, Québec
- **D^r Geoff Taylor**, Hôpital de l'Université de l'Alberta, Faculté de médecine, Division des maladies infectieuses, Edmonton, Alberta
- **D^{re} Mary Vearncombe**, directrice médicale, Prévention et contrôle des infections, Centre des sciences de la santé Sunnybrook, Toronto, Ontario

Les personnes suivantes représentaient l'Agence de la santé publique du Canada :

- **Toju Ogunremi**, BSc, MSC, analyste principale en recherche scientifique (chef de projet, ASPC)
- **Frédéric Bergeron**, IA, BScInf, MSc(A), infirmier-conseil
- **Katherine Defalco**, IA, BScInf, CIC, infirmière-conseil
- **Kathleen Dunn**, IA, BScInf, MScInf, gestionnaire
- **Jennifer Kruse**, IA, BScInf, infirmière-conseil
- **Laurie O'Neil**, IA, BScInf, infirmière-conseil
- **Shirley Paton**, IA, MScInf, conseillère technique principale
- **Christine Weir**, IA, BScInf, MS, CIC, infirmière épidémiologiste
- **D^r Tom Wong**, MD, MPH, FRCPC, directeur

BIBLIOGRAPHIE

1. Collaboration AGREE. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the Agree project, *Quality and Safety in Health Care*, vol. 12 (2003), p. 18-23.
2. Akobeng, A.K. Evidence based child health 3. Understanding Systematic Exams and meta-analysis, *Archives of Disease in Childhood*, vol. 90 (2005), p. 845-848.
3. Atkins, D., P.A. Briss et coll. Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system, *BioMed Central Health Services Research*, vol. 5 (2005) p. 25.
4. Bhandari, M., M.F. Swiontkowski, T.A. Einhorn, P. Tornetta, E. SchemSClch, P. Leece et J.G. Wright. Interobserver agreement in the application of levels of evidence to scientific papers, *American volume of the journal of bone and joint surgery (J. Bone Joint Surg. Am.)*, vol. 86 (2004), p. 1717-1720.
5. Campbell, F., H.O. Dickinson, J.V. Cook, F.R. Beyer, M. Eccles et J.M. Mason. Methods underpinning national clinical guidelines for hypertension: describing the evidence shortfall, *BMC Health Serv. Res.*, vol. 6 (2006), p. 47.
6. IRSC, CRSNG et CRSH. Énoncé de politique des trois Conseils : Éthique de la recherche avec des êtres humains, Chapitre 1-A : Recherche nécessitant une évaluation éthique : Règle 1.1, 1998 (avec modifications de 2000, de 2002 et de 2005). http://www.pre.ethics.gc.ca/policy-politique/tcps-epct/docs/TCPS%20October%202005_F.pdf
7. La Collaboration Cochrane. Cochrane handbook for systematic reviews of interventions. <http://www.cochrane.org/training/cochrane-handbook>
8. Ebell, M.H., J. Siwek, B.D. Weiss, S.H. Woolf, J.L. Susman, B. Ewigman et coll. Simplifying the language of evidence to improve patient care: Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in medical literature, *J. Fam. Pract.*, vol. 53, no 2 (févr. 2004), p. 111-120.
9. Flather, M.D., M.E. Farkouh et coll. Strength and limitations of meta-analysis: larger studies may be more reliable, *Controlled Clinical Trials*, vol. 18 (1997), p. 568-579.
10. Guyatt, G., D. Gutterman, M.H. Baumann, D. Addrizzo-Harris, E.M. Hylek, B. Phillips et coll. Grading strength of recommendations and quality of evidence in clinical guidelines: Report from an american college of chest physicians task force, *Chest*, vol. 129, no 1 (janv. 2006), p. 174-181. [voir commentaires]
11. Harbour, R. et J. Miller. A new system for grading recommendations in evidence based guidelines, *British Medical Journal*, vol. 323 (2001), p. 334-336.
12. Higgins, J.P.T., S.G. Thompson, J.J. Deeks et D.G. Altman. Measuring inconsistency in meta analyses, *BMJ*, vol. 327 (2003), p. 557-560.
13. Jackson, D. et E. Walters. Criteria for the systematic review of health promotion and public health interventions, *Health Promot Int.*, vol. 20 (2005), p. 367-374.

14. Liddle, J. et coll. *Method for evaluating research and guideline evidence (MERGE)*, Sydney New South Wales Department of Health, 1996.
15. Université Memorial de Terre-Neuve. *Research Requiring Ethics Review*.
<http://www.med.mun.ca/hic/Research%20Requiring%20Review.htm>
16. Moher, D., D.J. Cook et coll. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement, *Lancet*, vol. 354 (1999), p. 1 896–1 900.
17. Nordness, R. *Epidemiology and Biostatistics Secrets*, Philadelphie, Mosby Elsevier, 2006.
18. Norman, G.R. et D.L. Streiner. *Biostatistics: The Bare Essentials*. 2^e éd., B.C. Decker Hamilton-London, 2000.
19. Ramsay, C., L. Matowe et coll. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies, *Int J Technology Assessment in Health Care*, vol. 19 (2003), p. 613-623.
20. Rothman, K.J., S. Greenland et T.L. Lash, *Modern Epidemiology*, 3^e éd., Lippincott Williams & Wilkins, 2008.
21. Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developer's handbook, *SIGN*, Edinburgh, 2008.
22. Szklo, M. et F.J. Nieto. *Epidemiology: Beyond the basics*, Aspen Publishers, 2000.
23. Unal, B. et coll. Coronary heart disease policy models: a systematic review, *BioMed Central Public Health*, vol. 6 (2006), p. 213.