

Méthodes pour faire la synthèse d'études sans groupes témoins

Donna Fitzpatrick-Lewis • Donna Ciliska • Helen Thomas

Juin 2009



National Collaborating Centre
for Methods and Tools

Centre de collaboration nationale
des méthodes et outils

Méthodes pour faire la synthèse d'études sans groupes témoins

Préparé pour le Centre de collaboration nationale des méthodes et outils par
Donna Fitzpatrick-Lewis • Donna Ciliska • Helen Thomas

Juin 2009

Centre de collaboration nationale des méthodes et outils (CCNMO)

Université McMaster

1685, rue Main Ouest, bureau 302

Hamilton, Ontario L8S 1G5

Téléphone : 905.525.9140, poste 20455

Télécopieur : 905.529.4184

Web: www.ccnmo.ca

Financé par l'Agence de la santé publique du Canada

Affilié à l'Université McMaster

La production de ce document a été rendue possible grâce à l'aide financière de l'Agence de la santé publique du Canada. Les opinions exprimées ici ne représentent pas nécessairement celles de l'Agence de la santé publique du Canada.

Table des matières

Principales recommandations :	6
Sommaire	8
Introduction	12
Méthodes	13
Pertinence	13
Synthèse des données	14
Constatations	15
Outils d'évaluation critique	15
Examens systématiques	21
Méthodes pour faire la synthèse d'études quantitatives sans groupes témoins	28
Combiner des études qualitatives et quantitatives	31
Études comparant les résultats selon des méthodes différentes	32
Discussion	34
Recommandations	36
Tableau résumant les résultats des documents pertinents	38
Références bibliographiques	61

Principales recommandations :

1. Intégrer des études sans groupes témoins dans les examens systématiques, en particulier quand il n'y a aucune autre étude à prendre en compte. Les études sans groupes témoins peuvent aussi renseigner sur l'efficacité à long terme, les cas rares et les effets néfastes.
2. Éviter de généraliser l'orientation des écarts de revenu entre les études sur échantillon aléatoire et celles sur échantillon non contrôlé. Bien que les effets soient plus souvent de moindre envergure dans les essais sur échantillon aléatoire et contrôlé (EEAC), certains donnent des résultats inverses ou identiques quel que soit le modèle d'étude.
3. Adopter l'approche de Whitemore et Knafl (2005) selon laquelle il faut inclure les résultats d'essais sur échantillon non aléatoire. Whitemore et Knafl proposent une méthode d'examen par intégration en cinq étapes qui englobe la définition des problèmes, le dépouillement documentaire ainsi que l'évaluation, l'analyse et la présentation des données.
4. Recommandations concernant les méthodes d'examen systématique, incluant les études sans groupes témoins :
 - a) utiliser un outil d'évaluation critique dont l'uniformité interne, la fiabilité de test-retest, le coefficient d'objectivité et, à tout le moins, la validité apparente et de critère ont été éprouvés (Downs et Black, 1998; Slim, Nini, Forestier et coll., 2003; Steuten, Vrijhoef, van-Merode et coll., 2004; Thomas, Ciliska, Dobbins et coll., 2004a; Zaza, Wright-De Agüero, Briss et coll., 2000);
 - b) éviter de méta-analyser les résultats d'études par observation.
5. Suggestions aux auteurs d'études primaires :
 - a) assurer un modèle d'étude solide et transparent dans tous les essais sur échantillon non aléatoire;
 - b) confirmer que la question de recherche détermine le modèle d'étude;
 - c) inclure des renseignements détaillés sur l'étude; s'il manque de l'information, les lecteurs ne pourront déterminer ce qui a été fait ou non dans le modèle d'étude;
 - d) garantir l'uniformité de la terminologie et des descriptions dans toutes les études sur échantillon non aléatoire afin de faciliter les comparaisons.
6. Répercussions sur les recherches ultérieures :
 - a) effectuer des essais de fiabilité et de validité sur les outils d'évaluation critique;
 - b) mener d'autres études pour évaluer les divers résultats obtenus par des essais sur échantillon aléatoire ou non.
7. Il est difficile d'inclure des données provenant d'essais sur échantillon non aléatoire; cependant, cela peut offrir plusieurs avantages :
 - a) un ensemble de documentation plus large peut être inclus dans les straté-

- gies de recherche (Thomas, Harden, Oakley et coll., 2004);
- b) la portée des questions auxquelles on peut répondre peut être élargie;
 - c) l'étude permet d'établir pourquoi, comment et pour qui les interventions ou les programmes réussissent ou échouent (Oliver, Harden, Rees et coll., 2005).

Sommaire

Actuellement, il n'existe aucun modèle standard pour faire la synthèse des études sans groupes témoins. Le présent document a pour objet d'examiner les méthodes de synthèse, les outils d'évaluation critique et les études portant sur l'évaluation et la synthèse des études sans groupes témoins. Les EEAC sont réalisables en situation clinique, mais les interventions en santé publique peuvent rarement reproduire l'environnement contrôlé de la clinique. Les chercheurs, les responsables de l'élaboration des politiques et les décideurs en matière de pratique doivent souvent se fier à d'autres genres de modèles d'étude comme source de données probantes. Le présent document exposera dans leurs grandes lignes les précautions à prendre quand il s'agit d'inclure des études sur échantillon non contrôlé comme données probantes, et il recommande des outils efficaces pour analyser les résultats de ces genres d'études.

Méthodes

Avec l'aide d'un bibliothécaire, cinq bases de données électroniques ont été interrogées par mots clés. De plus, les références bibliographiques de tous les articles pertinents et de toute la littérature grise ont été dépouillées.

Un article répondait au critère de pertinence :

- s'il s'agissait d'une étude primaire comparant les résultats de la même intervention effectuée à l'aide d'EEAC ou d'études sans groupes témoins;
- s'il s'agissait d'une synthèse de méthodes employées pour examiner des études sans groupes témoins;
- si au moins il incluait l'une des méthodes suivantes : cohorte de groupes (avant-après ou juste après), modèle transversal, séries chronologiques interrompues (SCI) ou méthodes mélangées;
- s'il avait été publié entre 1995 et novembre 2007.

Un document était exclu :

- si l'accent était mis sur les EEAC, les essais cliniques sur échantillon contrôlé ou les études sur échantillon quasi-aléatoire;
- si l'accent était mis sur une étude analytique de cohortes (avant-après sur au moins deux groupes);
- si l'étude était seulement qualitative;
- s'il avait été publié avant 1995.

La mise à l'épreuve de la pertinence a été concluante dans le cas de trente-six documents, et ceux-ci ont fait l'objet d'une extraction de données complète. Bon nombre des études incluses étaient fondées sur l'observation. Il est important de noter que même si ces dernières études sont définies de bien des façons, seules celles qui avaient été menées sans groupes témoins ont été incluses dans le présent document.

Constatations

Quatorze articles fournissaient des outils d'évaluation critique pouvant aider à juger de la rigueur méthodologique des études primaires. Dans le cas d'un examen systématique, celles-ci devraient inclure une évaluation de leur qualité. Un outil d'évaluation critique est essentiel quand il s'agit de passer au crible les données probantes provenant d'études primaires sans groupes témoins. Ces outils peuvent renseigner sur les forces et les faiblesses de l'information. Ils aident le lecteur à évaluer la qualité méthodologique, à gérer les variables confusionnelles, à déterminer les biais possibles et à effectuer une analyse des données. Ils sont particulièrement utiles lorsque les critères d'évaluation des EEAC ne sont pas applicables. Dans bon nombre des cas, leur validité et leur fiabilité n'ont pas été mises à l'épreuve, ce qu'il serait sage de faire comme prochaine étape.

Onze examens systématiques comprenaient des études primaires sans groupes témoins, ou visaient à vérifier si les outils d'évaluation critique permettaient d'évaluer la qualité des études sans groupes témoins. Les examens destinés à vérifier s'il était possible de combiner les résultats d'EEAC et d'essais sur échantillon non aléatoire et contrôlé portent à croire que les études bien conçues peuvent donner des résultats semblables, peu importe leur genre. Bon nombre de ces examens portent à croire qu'une mauvaise réalisation de la méthode des études primaires nuit à la qualité plutôt qu'à la méthode elle-même. Lorsque l'on a tenté de comparer les données d'une étude à l'autre, il s'est avéré que la sous-communication des méthodes employées, la mauvaise gestion des variables confusionnelles et l'absence de terminologie analytique uniforme pouvaient être problématiques.

Six documents portaient sur diverses manières de faire la synthèse des données quantitatives découlant d'études sans groupes témoins et de les intégrer. Les auteurs ont convenu qu'il est avantageux d'intégrer des données probantes provenant de divers modèles d'étude, mais l'évaluation de la qualité est primordiale lorsqu'il s'agit de fournir des données probantes dignes de confiance. Les études primaires où le biais possible et l'impact ou l'effet probable sont clairs peuvent être intégrées dans les examens systématiques de manière utile.

Trois articles décrivaient les méthodes employées pour faire la synthèse de données provenant à la fois d'études qualitatives et quantitatives. Intégrer des données qualitatives et quantitatives dans un rapport peut contribuer à faire comprendre la réussite ou l'échec des interventions. Évaluer la qualité des études sans groupes témoins peut être problématique, surtout si la qualité est évaluée selon des critères traditionnels utilisés pour des examens systématiques ou des méta-analyses. Les résultats de l'évaluation de la qualité n'ont pas servi de critère d'exclusion, mais il pourra en être question lorsqu'il s'agira d'aborder les constatations. Il a été proposé de regrouper les constatations suivant une méthode thématique et globale comme autre approche appropriée à la méta-analyse. Les chercheurs pouvaient aussi intégrer certains des processus utilisés dans les examens systématiques, comme le fait de demander à deux personnes d'examiner et d'extraire les données. La plupart des auteurs ont suggéré une approche narrative à la présentation des résultats, quand la méta-analyse ne convenait pas.

Deux documents visaient à comparer les résultats de différentes méthodes appliquées à la même intervention. Ces articles soulignaient qu'il est utile de tenir compte de différents

modèles dans les examens. Non seulement il peut s'agir des meilleurs renseignements qui existent, mais ils permettent aussi de répondre aux questions concernant l'efficacité à long terme, les situations négatives et les cas rares. Les forces de l'inclusion de différents modèles dans les examens l'emportent sur les limites.

Recommandations

1. Intégrer des études sans groupes témoins dans les examens systématiques, en particulier quand il n'y a aucune autre étude à prendre en compte. Les études sans groupes témoins peuvent aussi renseigner sur l'efficacité à long terme, les cas rares et les effets néfastes.
2. Éviter de généraliser l'orientation des écarts de revenu entre les études sur échantillon aléatoire et celles sur échantillon non contrôlé. Bien que les effets soient plus souvent de moindre envergure dans les essais sur échantillon aléatoire et contrôlé (EEAC), certains donnent des résultats inverses ou identiques quel que soit le modèle d'étude.
3. Adopter l'approche de Whitemore et Knafl (2005) selon laquelle il faut inclure les résultats d'essais sur échantillon non aléatoire. Whitemore et Knafl proposent une méthode d'examen par intégration en cinq étapes qui englobe la définition des problèmes, le dépouillement documentaire ainsi que l'évaluation, l'analyse et la présentation des données.
4. Recommandations concernant les méthodes d'examen systématique, incluant les études sans groupes témoins :
 - a) utiliser un outil d'évaluation critique dont l'uniformité interne, la fiabilité de test-retest, le coefficient d'objectivité et, à tout le moins, la validité apparente et de critère ont été éprouvés (Downs et Black, 1998; Slim, Nini, Forestier et coll., 2003; Steuten, Vrijhoef, van-Merode et coll., 2004; Thomas, Ciliska, Dobbins et coll., 2004a; Zaza, Wright-De Agüero, Briss et coll., 2000);
 - b) éviter de méta-analyser les résultats d'études par observation.
5. Suggestions aux auteurs d'études primaires :
 - a) assurer un modèle d'étude solide et transparent dans tous les essais sur échantillon non aléatoire;
 - b) confirmer que la question de recherche détermine le modèle d'étude;
 - c) inclure des renseignements détaillés sur l'étude; s'il manque de l'information, les lecteurs ne pourront déterminer ce qui a été fait ou non dans le modèle d'étude;
 - d) garantir l'uniformité de la terminologie et des descriptions dans toutes les études sur échantillon non aléatoire afin de faciliter les comparaisons.
6. Répercussions sur les recherches ultérieures :
 - a) effectuer des essais de fiabilité et de validité sur les outils d'évaluation critique;

- b) mener d'autres études pour évaluer les divers résultats obtenus par des essais sur échantillon aléatoire ou non.
7. Il est difficile d'inclure des données provenant d'essais sur échantillon non aléatoire; cependant, cela peut offrir plusieurs avantages :
- a) un ensemble de documentation plus large peut être inclus dans les stratégies de recherche (Thomas, Harden, Oakley et coll., 2004);
 - b) la portée des questions auxquelles on peut répondre peut être élargie;
 - c) l'étude permet d'établir pourquoi, comment et pour qui les interventions ou les programmes réussissent ou échouent (Oliver, Harden, Rees et coll., 2005).

Introduction

Les politiques et la pratique en santé publique ont surtout pour but de protéger la santé des gens. Afin d'atteindre ce but, les professionnels de la santé publique tiennent compte des données probantes quand ils prennent des décisions. Lorsqu'il s'agissait d'analyser l'efficacité d'interventions, les scientifiques ont toujours préféré le modèle d'étude fondé sur des essais sur échantillon aléatoire et contrôlé (EEAC) pour obtenir les résultats les moins biaisés. Un EEAC peut procurer des résultats d'étude rigoureux du point de vue de l'efficacité. Cependant, ce ne serait peut-être pas le meilleur modèle d'étude à employer pour d'autres genres de questions de recherche (DiCenso, Prevost, Benefield et coll., 2004). En promotion de la santé publique et de la santé en général, il pourrait s'avérer difficile d'utiliser des essais sur échantillon aléatoire pour une question précise. Cela signifie-t-il que nous devrions laisser de côté les données probantes découlant d'autres méthodes? Certains chercheurs estiment que seuls les EEAC ont la rigueur scientifique nécessaire pour produire des données probantes valables.

En situation clinique, il est possible d'effectuer des EEAC. Mais les interventions en santé publique permettent rarement de reproduire l'environnement contrôlé de la clinique. Ces interventions sont souvent communautaires, et le recrutement peut s'avérer difficile, ce qui crée un biais de sélection. Par conséquent, il peut s'avérer impossible d'assurer la pureté des groupes témoins sans qu'ils soient intercontaminés. Les chercheurs en santé publique doivent souvent se fier à d'autres genres de modèles d'étude, qui occupent souvent un rang inférieur dans la « hiérarchie des données probantes » (Ogilvie, Egan, Hamilton et coll., 2005). Toutefois, les études par observation, par exemple, peuvent fournir des renseignements précieux lorsqu'il s'agit de prendre des décisions de santé publique. Les données qualitatives peuvent compléter les données statistiques en nous aidant à comprendre pourquoi et comment une intervention a réussi ou échoué et qui était en cause. Il existe toute une documentation sur les méthodes de synthèse des études qualitatives (méta-synthèses), mais ces méthodes dépassent la portée du présent document.

Étant donné que le risque de biais est plus grand dans les études par observation, celles-ci sont souvent exclues des examens systématiques de traitements. Toutefois, elles offrent plusieurs avantages par rapport aux EEAC : elles sont moins coûteuses, elles permettent un échantillonnage plus large, et elles procurent des mesures de résultats à plus long terme (Benson et Hartz, 2000). De plus, les études par observation peuvent fournir les « meilleures données probantes qui soient » dans certaines situations (Ogilvie et coll., 2005).

À l'heure actuelle, il n'existe aucun modèle standard pour faire la synthèse des résultats d'études sans groupes témoins. Le présent document a pour objet d'examiner les méthodes de synthèse, les outils d'évaluation critique et les études visant à évaluer les résultats d'études quantitatives sans groupes témoins et à en faire la synthèse. Il exposera aussi dans leurs grandes lignes certaines précautions à prendre lorsqu'il s'agit d'utiliser des études sur échantillon non contrôlé comme données probantes, et il présentera des recommandations de processus et d'outils.

Méthodes

Voici les méthodes qui ont été employées pour recueillir la documentation aux fins du présent document :

- un dépouillement exhaustif de la documentation publiée de janvier 1995 à octobre 2007;
- l'examen et la récupération de références bibliographiques tirées d'articles pertinents;
- le dépouillement et la récupération de littérature grise pouvant être pertinente.

La recherche initiale a été effectuée par un bibliothécaire qualifié, à l'aide des mots clés suivants : review (examen), systematic review (examen systématique), literature review (examen documentaire), meta-analysis (méta-analyse), not random (non aléatoire) et not randomized clinical trials (essais cliniques sur échantillon non aléatoire). Ces recherches par mots clés ont été effectuées dans plusieurs bases de données : PsycINFO, OVID MEDLINE, EMBASE, CINAHL et Scholars Portal. La recherche initiale a permis de trouver 9324 articles potentiellement pertinents. Les titres et les résumés ont été lus indépendamment par deux examinateurs, qui ont passé en revue 182 articles pour mettre pleinement leur pertinence à l'épreuve.

Pertinence

Un article faisait l'objet d'une mise à l'épreuve complète de sa pertinence :

- s'il s'agissait d'une étude primaire comparant les résultats de la même intervention effectuée à l'aide d'EEAC ou d'études sans groupes témoins;
- s'il s'agissait d'une synthèse de méthodes employées pour examiner des études sans groupes témoins;
- si au moins il incluait l'une des méthodes suivantes : cohorte de groupes (avant-après ou juste après), modèle transversal, séries chronologiques interrompues (SCI) ou méthodes mélangées;
- s'il avait été publié entre 1995 et novembre 2007.

Un document était exclu :

- si l'accent était mis sur les EEAC, les essais cliniques sur échantillon contrôlé ou les études sur échantillon quasi-aléatoire;
- si l'accent était mis sur une étude analytique de cohortes (avant-après sur au moins deux groupes);
- si l'étude était seulement qualitative;
- s'il avait été publié avant 1995.

La mise à l'épreuve de la pertinence a été concluante dans le cas de trente-six documents, et ceux-ci ont fait l'objet d'une extraction de données complète. Bon nombre des études incluses étaient fondées sur l'observation. Il est important de noter que même si ces dernières études sont définies de bien des façons, seules celles qui avaient été menées sans groupes

témoins ont été jugées pertinentes.

Synthèse des données

Les chercheurs ont fait la synthèse de 36 documents sous une forme narrative, par sujet :

- outils d'évaluation critique
- examens systématiques
- méthodes de synthèse des études quantitatives sans groupes témoins
- études qualitatives et quantitatives combinées
- comparaison de résultats suivant des méthodes différentes

Certains documents auraient pu être inclus dans plusieurs sections. Toutefois, nous avons classé les documents dans les groupes qui semblaient leur convenir le mieux.

Le document donne un résumé narratif des documents pertinents. Les résultats sont aussi exposés en détail sous forme de tableaux (reportez-vous à l'annexe 1).

Constatations

Outils d'évaluation critique

En tout, 14 documents décrivaient des outils d'évaluation critique.

Rangel, Kelsey, Colby et coll. (2003) se sont penchés sur l'absence d'un outil de mesure standard pour déterminer la qualité des études par observation. Ils ont mis au point un instrument d'évaluation de la qualité pour les études sur échantillon non aléatoire qui englobaient 30 éléments dans trois sous-échelles. Les sous-échelles servaient à évaluer la pertinence clinique, la méthode de communication et la force des conclusions. Les cotes globales ont été établies par combinaison de chaque sous-échelle, et les études étaient ainsi cotées comme étant « bonnes », assez bonnes » ou « mauvaises ». Le coefficient d'objectivité était élevé quand les degrés d'entente atteignaient une concordance de résultats de 84,6 % (n = 1573 éléments). Pour l'ensemble des sous-échelles, il allait de 73,3 % à 85,8 % (n = 60 à 1258). Voici les applications possibles de cet outil :

- favoriser la mise au point d'une méthode standard pour les examens systématiques de données rétrospectives;
- aider à créer une stratégie de méta-analyse des essais sur échantillon non aléatoire;
- élaborer des lignes directrices standard de communication pour les revues examinées par les pairs.

Margetts et coll. (1995) ont mis au point un système de notation pour juger de la qualité scientifique des études épidémiologiques par observation établissant un lien entre l'alimentation et le risque de cancer. Les études cas-témoin et de cohortes ont été examinées séparément, car certains des repères de fiabilité s'appliquaient seulement à l'un de ces genres d'études. Les études cas-témoin ont été notées dans trois grands domaines : la qualité de l'évaluation alimentaire, le recrutement des sujets et l'analyse des résultats. Les études de cohortes ont été notées dans quatre domaines : l'évaluation alimentaire, la définition des cohortes, la confirmation des cas de maladie et l'analyse des résultats. Le coefficient d'objectivité a été évalué et a donné lieu à un degré d'accord élevé entre les examinateurs, les examinateurs de cohortes s'étant entendus légèrement mieux que les autres. Ce prototype de système de notation a permis aux auteurs de décrire les données épidémiologiques sur l'alimentation et le cancer. Cependant, il se peut qu'il ne puisse être appliqué à d'autres sujets.

Downs et Black (1998) ont dressé et mis à l'épreuve une liste de contrôle pour évaluer les études sur échantillon aléatoire ou non aléatoire. Pour la créer, ils se sont servis de principes d'épidémiologie, d'examens de modèles d'étude et de listes de contrôle existantes. La liste est composée de 26 éléments répartis en cinq sous-échelles :

- communication
- validité externe
- biais

- confusion
- puissance

La note maximum a été de 31. Vingt-trois des questions auraient pu être posées pour n'importe quelle étude sur n'importe quelle intervention de soins de santé. Les trois autres étaient propres à un sujet et destinées à fournir aux évaluateurs des renseignements sur les variables confusionnelles, les résultats principaux et la taille d'échantillon requise pour qu'elle revête une importance clinique et statistique ($p < 0,05$). L'indice de qualité était très uniforme à l'interne, avait une bonne fiabilité de test-retest et un bon coefficient d'objectivité et était très valable tant en apparence que suivant les critères. La liste était aussi efficace que les autres listes de contrôle établies pour les essais sur échantillon aléatoire, et il y avait peu de différence quant à son efficacité avec les études sur échantillon aléatoire ou non.

Afin de s'attacher au manque d'instruments validés pour mesurer la qualité des essais par observation ou sur échantillon non aléatoire, Slim et coll. (2003) ont mis à l'essai un indice méthodologique pour les études sur échantillon non aléatoire (MINORS). Cet outil comprend 12 points (les huit premiers concernent spécifiquement les essais non comparatifs) :

- un objectif clairement énoncé
- l'inclusion de patients consécutifs
- une collecte prospective des données
- des critères convenant à l'objectif de l'étude
- une évaluation impartiale des critères de l'étude
- une période de suivi appropriée à l'objectif de l'étude
- un suivi impossible dans moins de 5 % des cas
- un calcul prospectif de la taille de l'étude
- un groupe témoin adéquat
- des groupes contemporains
- une équivalence de base des groupes
- une analyse statistique adéquate

Cet outil offrait une bonne fiabilité, une bonne uniformité interne et une bonne validité.

Zaza et coll. (2000) ont créé un format pour classer et fournir des éléments descriptifs d'inventions en santé publique et pour évaluer la qualité d'exécution de l'étude. Les auteurs ont reconnu que tous les modèles d'étude présentent des difficultés propres. Par conséquent, les questions de qualité ont été élaborées en fonction de celles-ci. Par exemple, les auteurs ont laissé entendre que même si le traitement à l'aveugle est un élément important à considérer dans le cas des essais sur échantillon aléatoire, il n'est pas approprié d'évaluer la validité suivant un modèle à séries chronologiques. Zaza et coll. ont donné des catégories de questions pour évaluer les menaces possibles à la validité des résultats dans les études primaires, y compris :

- la description de l'invention
- l'échantillonnage

- la fiabilité et la validité des mesures de résultat
- l'opportunité de l'analyse statistique
- l'interprétation des résultats

Leur approche avait notamment pour objectif de concevoir un outil assez flexible pour pouvoir servir à évaluer des modèles d'étude et des interventions divergents. Le formulaire et la procédure standard de résumé ont amélioré la validité et la fiabilité du Guide to Community Prevention Services: Systematic Reviews and Evidence-Based Recommendations (guide des services communautaires de prévention : examens systématiques et recommandations fondées sur des données probantes).

Greer et coll. (2000) ont présenté et abordé un système de notation des données probantes qui a été mis au point par l'Institute for Clinical Systems Improvement (ICSI). Cette approche a été créée à la suite d'une rétroaction selon laquelle un système existant ne fonctionnait pas. Le système a pour but d'aider les professionnels occupés qui doivent juger de la qualité de données probantes. Après avoir examiné bien d'autres systèmes de notation, le groupe de travail de l'ICSI a convenu qu'il importait de séparer l'évaluation des rapports de recherche et de celle de l'ensemble des données probantes appuyant la conclusion. La feuille de calcul subséquente pour la notation des données probantes (l'outil primaire) comprenait un énoncé de conclusion, un résumé des rapports de recherche qui appuient ou mettent en doute la conclusion, l'attribution des catégories et des indicateurs de qualité aux rapports de recherche et l'attribution d'une note à la conclusion. Les rapports primaires de nouvelle collecte de données se sont vus attribuer une note à lettre de A à D, selon le genre d'étude : EEAC, étude de cohortes, essai sur échantillon non aléatoire avec groupes témoins concurrents ou historiques, étude cas-témoin, étude de sensibilité et de spécificité d'un essai diagnostic, étude descriptive fondée sur la population, étude transversale, séries de cas ou rapport de cas. Le rapport comprenait aussi un système de qualité en cinq points pour coter les attributs positifs et négatifs des modèles d'étude, où les cotes étaient positives, négatives ou neutres. Voici quelles étaient les questions :

- critère d'inclusion ou d'exclusion a priori
- biais
- résultats statistiquement ou cliniquement significatifs
- applicabilité des résultats à d'autres populations
- modèle d'étude clairement établi

Le système a été appliqué à plus de 40 lignes directrices et rapports d'évaluation technologique de l'ICSI. Les auteurs ont signalé qu'il semble réussir à réduire la complexité d'autres systèmes de notation, bien qu'il produise une classification défendable des conclusions fondée sur la force des données probantes sous-jacentes.

Steuten et coll. (2004) ont critiqué les outils utilisés pour évaluer la qualité méthodologique de l'évaluation des technologies de la santé (ÉTS) pour la gestion des maladies. Les auteurs ont dressé une liste de problèmes qui surviennent lorsqu'il s'agit d'évaluer des études d'ÉTS, y compris : le modèle d'étude (EEAC ou essai par observation), les critères de sélection et de limitation des patients, les mesures de base et de résultat, le traitement à l'aveugle des patients et des fournisseurs, la description des interventions complexes à fac-

ettes multiples, et l'évitement des co-interventions. Ils ont créé, proposé et validé un nouvel instrument pour évaluer la qualité méthodologique de l'ÉTS pour la gestion des maladies qui comprend quatre éléments :

- population à l'étude
- description de l'intervention
- mesure des résultats
- analyse ou présentation des données

L'instrument compte 15 éléments qui englobent la validité interne, la validité externe et les considérations statistiques. Il est fiable du point de vue du classement hiérarchique, de la fiabilité de test-retest et du coefficient d'objectivité lorsqu'il est question des ÉTS pour la gestion des maladies.

Deux rapports (Chou et Helfand, 2005; McIntosh, Woolacott et Bagnall, 2004) ont laissé entendre que les preuves de préjudice sont aussi importantes que celles d'effet. Cependant, les examens systématiques standard axés sur l'efficacité et les essais sur échantillon aléatoire peuvent ne pas représenter un modèle efficace d'évaluation des préjudices. La documentation sur les préjudices se trouve dans des essais non publiés, des études par observation et de la littérature grise. Appliquer des listes de contrôle de la qualité actuelles à cet ensemble de données s'est avéré problématique et ne permettait pas volontiers de rendre les genres d'information qui concernaient les effets nocifs. McIntosh et coll. ont adapté des listes de contrôle publiées pour traduire l'information sur les effets nocifs qui se trouvait dans les études de cohortes par observation. Cette nouvelle liste de contrôle comprenait aussi des éléments comme comment et quand les événements ont été signalés, et si le moment où ils ont eu lieu pendant l'étude a été enregistré. Toutefois, le pire obstacle à l'application de toute liste de contrôle a été le manque de détails méthodologiques fournis par les auteurs des études primaires. Les chercheurs ont aussi inclus des états narratifs des résultats. Chou et Helfand (2005) ont examiné des comptes rendus de cas et des études par observation sur les effets nocifs de traitements. Ils ont aussi éprouvé des difficultés à évaluer la qualité, car les instruments qui auraient pu servir à évaluer ces genres d'études étaient rares ou incohérents. La rigueur d'élaboration, la portée et le nombre et le genre d'éléments utilisés étaient tous des questions litigieuses. Ils ont procédé à des essais pilotes sur un outil d'évaluation de la qualité pour les EEAC, les essais cliniques sur échantillon contrôlé et les études de cohortes dans le cas d'une série d'interventions chirurgicales non contrôlées où l'on signalait des complications découlant de l'endartériectomie de la carotide. Cet outil en huit points indiquait bon, assez bon ou mauvais. L'outil était mis à l'épreuve sur des études concernant une intervention. Les auteurs ont prévenu les chercheurs d'éviter le regroupement inapproprié de résultats statistiques provenant d'études par observation, étant donné le risque de biais dû au mauvais contrôle des variables confusionnelles et de biais de sélection.

Le GRADE Working Group (Atkins, Briss, Eccles et coll., 2005) a fait état de la mise au point et de l'essai pilote de l'approche GRADE à l'évaluation des données probantes et des recommandations. Douze profils de données probantes ont été cotés indépendamment par 17 juges qui étaient tous versés en utilisation d'autres approches à l'évaluation de données probantes et de recommandations. Chaque profil de données probantes reposait sur les renseignements fournis dans un examen systématique et comprenait deux tableaux,

un pour l'évaluation de la qualité et un pour le résumé des constatations. Le processus d'évaluation de la qualité était conçu pour que chaque résultat soit évalué séparément. Le tableau des résultats présentait le nombre d'études où figurait le résultat en question, le modèle d'étude (EEAC ou par observation) et la qualité des études. L'évaluation de la qualité des résultats a révélé que le jugement de l'examineur sur la qualité était influencé par d'autres facteurs que le modèle d'étude, la qualité, l'uniformité et la franchise. Ces autres facteurs étaient les données éparses, les solides associations, les biais de publication, la réaction aux doses et les situations où toutes les variables confusionnelles plausibles renforçaient la confiance envers l'orientation de l'effet. Les juges devaient indiquer la facilité de compréhension et la sensibilité de l'approche; ils ont généralement convenu que l'approche GRADE était claire, compréhensible et sensée.

Khan et coll. (2001) ont fourni des renseignements sur ce qu'il faut inclure dans toute liste de contrôle pour l'évaluation de la qualité. Dans le cas d'essais sur l'efficacité, il faudrait d'abord songer à des EEAC quand ils existent et sont bien conçus. En l'absence de bons EEAC, il y aurait lieu d'envisager des études quasi-expérimentales et par observation (QEO) bien conçues. Khan et coll. ont fourni plusieurs questions à étudier lorsqu'il s'agit d'évaluer la qualité d'études par observation. Ils ont souligné que même s'il est possible de tirer des conclusions de ces études, on ignore peut-être au juste si les groupes étudiés sont comparables. Les auteurs ont laissé entendre qu'il faut rapporter les résultats avec précaution, souvent en recommandant des recherches plus approfondies.

Thomas, Ciliska, Dobbins et Micucci (2004) se sont servis d'un instrument d'évaluation de la qualité qui permettait de coter du point de vue méthodologique des études qui n'étaient pas des EEAC, en vue de leur inclusion dans un examen systématique. Cette méthode était précisément destinée à évaluer la qualité des recherches en santé publique. Voici quels étaient les éléments d'évaluation :

- biais de sélection
- modèle d'étude
- variables confusionnelles
- traitement à l'aveugle
- méthodes de collecte des données
- retraits ou abandons

Les études cotées par ce système obtenaient une cote méthodologique forte, moyenne ou faible. L'instrument avait un bon coefficient d'objectivité et une bonne fiabilité de test-retest. Il avait aussi prouvé la validité du contenu et des critères; les EEAC et les ECEC étaient jugés solides, et les essais sur échantillon non aléatoire et contrôlé (EENC), comme les études de cohortes et par observation, ont été jugés moyens. Cependant, les études sur échantillon non aléatoire pouvaient atteindre une cote globale moyenne si les autres éléments de l'étude étaient solides. Les données sont extraites à même les études et sont rapportées sous forme narrative. La méta-analyse était rarement effectuée, car il s'était avéré que les populations et les interventions visées par les études en santé publique sont souvent trop hétérogènes pour ce genre d'analyse. L'extraction des données a permis de rapporter des effets (ou un manque d'effets) d'une grande portée statistique de chaque étude, et les chercheurs ont émis des commentaires sur l'importance de tout effet statistiquement

significatif.

Ramsay et coll. (2003) ont examiné la qualité des séries chronologiques interrompues (SCI) à l'aide d'études incluses dans deux examens systématiques. Ils ont mis au point un outil d'évaluation de la qualité suivant les critères suivants :

- l'intervention se déroulait indépendamment d'autres changements qui se produisaient au fil du temps;
- l'intervention était peu susceptible de toucher la collecte des données;
- le résultat était évalué à l'aveugle ou mesuré objectivement;
- le résultat était fiable ou mesuré objectivement;
- la composition de l'ensemble de données à chaque point abordé englobait au moins 80 % de tous les participants à l'étude;
- la forme de l'effet d'intervention était précisée au préalable;
- le raisonnement sous-tendant le nombre et l'espacement des points de données était décrit;
- l'étude était analysée convenablement selon des techniques par séries chronologiques.

Ils ont constaté que 37 des 58 études avaient été mal analysées, et que 33 d'entre elles devaient être analysées à nouveau. De mauvais modèles d'étude, une puissance insuffisante et une analyse inappropriée des études par SCI ont donné lieu à la présentation de conclusions trompeuses dans la documentation publiée. Les auteurs ont recommandé que toutes les données provenant d'études par SCI soient analysées à nouveau avant d'être incluses dans les examens.

Conn et Rantz (2003) ont examiné des moyens par lesquels les chercheurs peuvent gérer la qualité lorsqu'ils envisagent des études primaires dans le cadre d'une méta-analyse. Bien qu'il existe plus de 100 échelles pour mesurer la qualité de ces études, elles peuvent varier selon la taille, la composition, la complexité et l'étendue. Établir la validité des échelles est compliqué et difficile. Les échelles ne sont pas fiables et précises dans tous les domaines scientifiques, et leur application peut s'avérer problématique et inégale. Conn et Rantz ont exposé dans leurs grandes lignes trois méthodes qui permettraient de gérer la qualité :

- fixer un seuil de qualité;
- pondérer par la cotation de la qualité;
- considérer la qualité comme question empirique.

Cependant, comme méthode autonome, chacune a ses limites. Par exemple, fixer un seuil de qualité peut donner lieu à l'exclusion de travaux de recherche moins rigoureux (où des données peuvent renfermer des renseignements importants). La pondération par la qualité d'étude peut être limitée par les outils d'échelonnement utilisés et certains problèmes potentiellement inhérents, comme le coefficient d'objectivité. Enfin, considérer la qualité comme une question empirique peut mener à des mesures globales masquant des effets intéressants d'éléments individuels de qualité sur des estimations d'ampleur de l'effet. Afin de tenir compte de ces limites, les chercheurs pourront devoir combiner des stratégies dans le but d'inclure des études jugées méthodologiquement faibles, mais contenant quand même des

renseignements importants.

Résumé de la section

Examiner la documentation à inclure dans un examen systématique devrait inclure une évaluation de la qualité. Un outil d'évaluation critique est essentiel lorsqu'il s'agit de passer en revue les données probantes dans des études primaires sans groupes témoins. Ces outils peuvent fournir des renseignements sur les forces et les faiblesses des renseignements. Les outils d'évaluation critique aident le lecteur à évaluer les données probantes fournies d'après la qualité méthodologique, la gestion des variables confusionnelles, les biais possibles et l'analyse des données. Ces outils sont particulièrement utiles quand les systèmes d'évaluation employés pour les EEAC ne sont pas applicables. La validité et la fiabilité de bon nombre des outils indiqués dans les présentes n'ont pas été éprouvées. Ce serait raisonnable de le faire comme prochaine étape.

Examens systématiques

Onze examens systématiques portaient sur la synthèse de données provenant d'études sans groupes témoins.

Linde, Scholz, Melchart et Willich (2002) ont cherché à déterminer si les examens systématiques devaient inclure à la fois des études sur échantillon aléatoire et non aléatoire. Les chercheurs ont examiné les données sur le recours à l'acupuncture pour les maux de tête chroniques afin d'étudier les questions suivantes :

1. Les études sur échantillon aléatoire et non aléatoire sur l'acupuncture pour les maux de tête chroniques diffèrent-elles du point de vue des patients, des interventions, des aspects qualité indépendants de la conception et des taux de réponse?
2. Les études sur échantillon non aléatoire fournissent-elles d'autres renseignements pertinents (résultats à long terme, facteurs prévisionnels, effets néfastes ou complications et taux de réponse chez les groupes représentatifs et bien définis de patients)?
3. Si les taux de réponse chez les patients choisis de manière aléatoire ou non diffèrent, quelles sont les explications possibles?

Cinquante-neuf études répondaient aux critères d'inclusion, et 24 d'entre elles étaient des essais sur échantillon aléatoire et 35, des essais sur échantillon non aléatoire (cinq essais sur échantillon non aléatoire et contrôlé, 10 essais prospectifs sur échantillon non contrôlé, 10 séries de cas et 10 enquêtes transversales). En tout, 535 patients ont reçu un traitement à l'acupuncture dans les 24 EEAC, comparativement à 2695 dans les 35 études sur échantillon non aléatoire. En moyenne, les essais sur échantillon aléatoire étaient faits sur un échantillon de moindre taille, respectaient davantage les critères de qualité et offraient des taux de réponse plus faibles que le traitement (0,59; 0,40 – 0,69) c. (0,78; 0,72 – 0,83). Peu importe l'état de randomisation, dans les études qui répondaient davantage aux critères de qualité, les taux de réponse étaient plus faibles que les taux de traitement. La période de suivi n'était pas nettement plus longue dans les études sur échantillon non aléatoire. Dans le cas de l'acupuncture appliquée aux maux de tête chroniques, les études sur échantillon non aléatoire confirmaient les constatations tirées d'une étude antérieure (Greenhalgh, Robert,

Macfarlane et coll., 2005) selon lesquelles le traitement était susceptible d'être efficace. Cependant, les études sur échantillon non aléatoire procuraient peu de renseignements supplémentaires pertinents sur les effets à long terme, les facteurs prévisionnels ou les effets négatifs. En général, les auteurs concluaient que les études sur échantillon non aléatoire de bonne qualité donnaient des résultats semblables à ceux des EEAC. Cela porte à croire que leur inclusion dans les examens systématiques, bien qu'elle augmentait la charge de travail des auteurs desdits examens, permet un élargissement utile aux fins de généralisabilité.

MacLehose et coll. (2000) ont enquêté sur le lien entre la qualité méthodologique et les estimations d'efficacité en comparant les EEAC aux études quasi-expérimentales et par observation (QEO). Pour examiner la documentation, les auteurs ont employé deux stratégies :

1. comparer les estimations quant à l'efficacité de toute intervention qui ont été obtenues par des EEAC et des études QEO, où les deux estimations figuraient dans un seul document;
2. comparer les estimations quant à l'efficacité d'interventions précises qui ont été obtenues par des EEAC et des études QEO, où les deux estimations figuraient dans des documents différents.

Dans le cas de la 1^{re} stratégie, les auteurs ont relevé 14 documents qui contenaient 38 comparaisons, dont 13 étaient jugées de qualité supérieure et 25, de faible qualité. Les critères d'évaluation de la qualité étaient les suivants :

- les estimations provenaient de la même population;
- le traitement à l'aveugle des évaluateurs de résultat;
- la mesure dans laquelle l'estimation obtenue par l'étude QEO a été établie compte tenu du risque de variables confusionnelles.

Selon les constatations, les écarts entre les estimations d'ampleur de l'effet et de la fréquence de résultat qui ont été obtenues par des EEAC et des études QEO pour les groupes ayant fait l'objet d'une intervention et les groupes témoins étaient plus faibles pour les comparaisons de qualité supérieure que pour celles de faible qualité. Par ailleurs, les estimations de taille d'ampleur obtenues par des études QEO n'avaient pas tendance à être extrêmes pour les comparaisons de qualité supérieure que pour les EEAC de faible qualité. Dans le cas de la 2^e stratégie, les interventions spécifiques comprenaient le dépistage par mammographie destiné à réduire la mortalité due au cancer du sein et les suppléments d'acide folique servant à réduire les anomalies du tube neural. Les auteurs ont relevé 34 documents, dont 12 étaient des EEAC, 11, des essais sur échantillon non aléatoire ou des études de cohortes et neuf, des études cas-témoin appariées ou non. Ces études ont été évaluées d'après :

- la qualité des rapports;
- la généralisabilité des résultats;
- la mesure dans laquelle les estimations d'efficacité auraient pu être influencées par un biais ou une variable confusionnelle.

Les notes de qualité étaient plus élevées pour les études de cohortes et cas-témoin que pour les EEAC, et elles étaient moins élevées pour les études de cohortes que pour les

études cas-témoin. La méta-régression des attributs d'étude par rapport aux estimations du risque relatif n'a révélé aucun lien entre l'ampleur de l'effet et la qualité de l'étude. Les estimations tirées des EEAC et des études de cohortes n'étaient pas nettement différentes. Toutefois, les études cas-témoin ont donné des estimations bien différentes lorsqu'il était question de mammographie (avantage supérieur) et d'acide folique (avantage inférieur).

Britton et coll. (1998) ont étudié des questions touchant le processus de randomisation qui peuvent influencer sur la validité des conclusions tirées des résultats d'EEAC et d'études sur échantillon non aléatoire (y compris des études sans groupes témoins). Les auteurs ont posé quatre questions de recherche pour examiner la randomisation et la validité :

1. Les études sur échantillon non aléatoire diffèrent-elles systématiquement des EEAC du point de vue de l'effet du traitement?
2. Existe-t-il des différences systématiques entre les personnes incluses et exclues, et influencent-elles l'effet mesuré du traitement?
3. Dans quelle mesure est-il possible d'adapter en fonction des différences de base entre les groupes à l'étude?
4. En quoi la préférence des patients est-elle importante sur le plan des résultats?

Leur étude englobait 18 documents où l'on comparait directement les résultats d'EEAC et d'études sur échantillon non aléatoire. Les résultats n'ont révélé aucun compte rendu uniforme d'estimations importantes ou moindres de l'effet du traitement, fondées sur le modèle d'étude. Les auteurs ont aussi rapporté que le genre d'intervention ne semblait pas exercer d'influence; cependant, ils ont lancé un avertissement selon lequel il faudrait étudier davantage afin de valider cette conclusion. Dans les EEAC, les exclusions globales étaient courantes, et la proportion de sujets admissibles inclus allait de 1 % à 100 %. Les grandes bases de données cliniques renfermant des renseignements détaillés sur la gravité des cas et le pronostic ont remplacé les EEAC. Lorsque les sujets tirés des bases de données étaient choisis selon les mêmes critères d'inclusion que pour les EEAC, les effets du traitement des deux modèles étaient semblables. La documentation des caractéristiques des personnes admissibles qui n'ont pas participé aux essais était mal inscrite dans la plupart des EEAC. Dans les EEAC, il se peut que les mesures d'effets de traitement soient exagérées en raison de la participation plus large des universités et des centres d'enseignement, comparativement aux études sur échantillon non aléatoire. Dans celles-ci, le redressement en fonction des différences a souvent modifié l'ampleur de l'effet du traitement, mais pas nettement. Mais surtout, l'orientation du changement était constante. Seulement quatre documents s'attachaient au rôle de la préférence du patient dans les résultats. Dans ces documents, la préférence représentait une partie des différences observées entre les modèles d'étude. Britton et coll. ont conclu ce qui suit :

- une étude sur échantillon non aléatoire bien conçue est préférable à une petite EEAC mal conçue et exclusive;
- les EEAC devraient inclure une foule de cadres d'exercice;
- les populations étudiées devraient être représentatives de tous les patients qui font l'objet de l'intervention;
- les exclusions pour des raisons de commodité administrative devraient être rejetées;

- il faudrait réduire les différences au minimum pour s'assurer que les sujets des deux genres d'étude sont comparables.

Lemmer, Grellier et Stevens (1999) ont modifié le Cochrane Collaboration Protocol for Systematic Reviews pour un rapport qui portait sur les données probantes concernant la prise de décisions et les visites médicales (fourniture de soins de santé à domicile) en Grande-Bretagne. Les visites médicales et la prise de décisions sont influencées par des facteurs comme les forces sociales et environnementales, qui tendent à ne pas être saisies dans les données probantes provenant d'EEAC. Dans le Cochrane Protocol adapté, on a omis les sections jugées inappropriées pour des documents de recherche fondés sur des EENC. La rigueur méthodologique et la pertinence de chaque article ont été établies suivant une échelle de notation numérique allant de 8 (maximum) à 1 (minimum). Les auteurs n'ont pas fourni les variables utilisées pour mesurer la rigueur méthodologique. Deux membres de l'équipe de recherche ont examiné chaque article, et des réunions de concertation hebdomadaires ont été tenues pour s'entendre sur les différences de notation. Les notes devaient donner un aperçu de la méthodologie et de la pertinence de chaque article. Cependant, la note globale n'indiquait nullement la répartition des notes. Les examinateurs se sont prononcés sur l'opportunité et l'importance de chaque article, mais leurs commentaires ne faisaient pas partie de la notation. Les auteurs ont constaté que le système de notation constituait un moyen utile d'orienter la discussion aux réunions de concertation. La notation a été influencée par l'interprétation qu'ont faite les examinateurs des méthodologies qualitatives et la pertinence des articles. Ils ont aussi laissé entendre que cette approche plus ouverte a permis d'examiner des articles qui renfermaient des sections ou des renseignements importants malgré les faibles notes accordées. Certains documents contenaient de très brèves sections sur les méthodes, ce qui a rendu la notation difficile et a donné de faibles notes.

Reconnaissant qu'il est difficile de faire la synthèse des diverses données fournies dans les examens, Goldsmith, Bankhead et Austoker (2007) ont mis au point une nouvelle méthode d'examen. Ils l'ont appliquée à un examen de données probantes pour influencer des lignes directrices concernant le contenu des renseignements sur les patientes relativement au dépistage du cancer du col de l'utérus. Ils ont suivi des procédures standard afin de mener un examen systématique, y compris :

- la conception d'études a priori
- un dépouillement documentaire exhaustif
- l'examen des titres et des résumés par deux personnes indépendantes
- l'évaluation de la qualité
- l'extraction des données

Les chercheurs se sont servis de listes de contrôle pour évaluer la qualité des études quantitatives et qualitatives. Des listes de contrôle établies ont été utilisées dans le cas d'études quantitatives (Critical Appraisals Skills Program (CASP), Scottish Intercollegiate Guidelines Network (SIGN), New Zealand Guidelines Group (NZGG) et UK Government Chief Social Researcher's Office (UKGCSRO)) (CASP, 2005; Khan et coll., 2001; Lethaby, Wells et Furness, 2001; Spencer, Ritchie et Lewis, 2004). Les chercheurs ont dressé une liste de contrôle pour les études qualitatives qui comprenait l'objet, la population, le taux de réponse, la définition des résultats et l'évaluation. Toutes les listes de contrôle comprenaient une sec-

tion commentaires. La qualité méthodologique des études était cotée ++ (tous les critères ou presque ont été remplis), + (certains critères ont été remplis) ou – (aucun critère n'a été rempli ou peu d'entre eux l'ont été). L'évaluation de la qualité a donné une idée des forces et des faiblesses de chaque étude. Cependant, les failles méthodologiques n'ont pas entraîné l'exclusion d'études. Les notes de qualité ont été intégrées pendant la phase de synthèse des données de l'examen. Goldsmith et coll. (2005) ont intégré les directives de synthèse établies dans le système GRADING (Atkins et coll., 2005) abordé dans la section Outils d'évaluation critique du présent document.

DeWalt et coll. (2004) a mené un examen systématique des études qui a permis de mesurer l'alphabétisation et un ou plusieurs résultats pour la santé. Les modèles d'étude admissibles étaient par observation : études de cohortes prospectives et rétrospectives, études cas-témoins et études transversales. Leur examen s'est fait selon un processus standard d'examen systématique, y compris :

- le dépouillement documentaire exhaustif de plusieurs bases de données électroniques,
- des questions de recherche bien définies,
- des critères d'inclusion et d'exclusion,
- l'examen d'articles complets par deux auteurs,
- l'évaluation de la qualité et le compte rendu des constatations.

Les chercheurs ont adapté l'outil d'évaluation de la qualité d'après West et coll. (2002). Chaque étude a été cotée selon les critères suivants :

- la population à l'étude
- la comparabilité des sujets
- la validité et la fiabilité de la mesure d'alphabétisation
- le maintien de groupes comparables
- la mesure des résultats
- l'analyse statistique
- le contrôle des variables confusionnelles

D'après ces critères, les études ont été cotées comme étant bonnes, assez bonnes ou mauvaises. L'outil n'a pas été validé. Bien que les auteurs de l'examen aient pu fournir une analyse statistique des résultats pour l'alphabétisation et la santé, ils ont précisé que les données d'études primaires comportaient bien des défis en raison des diverses mesures de lecture utilisées et des points de coupure aux fins d'analyse. De plus, le manque de mesures statistiques, le contrôle inadéquat des variables confusionnelles et le manque d'adaptation aux fins de comparaisons multiples ont rendu difficiles les comparaisons entre études.

Thomson et coll. (2006) ont examiné la communication des données sur les déterminants socio-économiques de la santé au Royaume-Uni, afin d'établir si les investissements de l'État dans le domaine avaient amélioré la santé de la population. Les données de 10 évaluations ont été synthétisées suivant des méthodes d'examen systématique, y compris :

- une stratégie de recherche,
- des critères d'inclusion ou d'exclusion a priori,
- l'examen d'articles par deux personnes indépendantes,
- l'extraction des données,
- la synthèse des données.

Huit des évaluations d'impact reposaient sur des études de cas où les données provenaient de quelques sites pour que le programme national soit représenté. Les questions méthodologiques relatives aux rapports d'évaluation englobaient les mauvaises méthodes d'évaluation et sources de données et les échantillons de faible taille. Les auteurs ont recouru à une synthèse narrative pour faire état des constatations. Ils ont signalé que cette approche constituait un nouveau moyen d'établir des données probantes, le message adapté pour qu'il influe sur l'élaboration des politiques publiques en santé. La synthèse était mise en question par les lacunes méthodologiques des études incluses.

Deux études (Stein, Dalziel, Garside et coll., 2005; Dalziel, Round, Stein et coll., 2005) portaient sur la qualité des données probantes sur l'efficacité utilisées dans les ÉTS. Les séries de cas, communément employées dans les ÉTS, constituent une forme faible dans la hiérarchie des données probantes. Idéalement, les données de séries de cas ne serviraient pas à des examens systématiques; cependant, il arrive qu'elles soient les seules données probantes qui existent. Ne trouvant aucune étude semblable dans leur dépouillement documentaire, Stein et coll. ont examiné un échantillon d'examens systématiques de séries de cas portant sur le lien entre les caractéristiques des séries de cas et le résultat. Ils n'ont pas comparé les résultats de séries de cas avec ceux d'EEAC. Dans leur analyse, ils ont découvert peu de données probantes sur le lien entre les caractéristiques méthodologiques, la taille de l'échantillon ou l'approche prospective et le résultat. Ils ont fourni un rapport narratif de leurs constatations et ont inclus un tableau de résultats individuels des variables analysées. Ils ont précisé que toutes les interventions étudiées étaient chirurgicales, ce qui pouvait limiter la généralisabilité de leurs constatations. Dans une deuxième étude, Dalziel et coll. ont comparé les résultats d'EEAC et d'études de séries de cas pour des interventions chirurgicales. Ils ont examiné des rapports qui incluaient des données provenant d'EEAC et de séries de cas. Ils ont comparé ces études à l'aide de la partie intervention des EEAC comme élément de comparaison : la méta-analyse d'EEAC a été comparée à de solides régressions pondérées en utilisant l'intervention comme facteur confusionnel et en évaluant la grandeur de coefficient. Leur analyse a révélé que les EEAC ne montraient aucune différence de résultat entre les types de traitement, alors que les séries de cas montraient une augmentation de la mortalité de 1 à 2 % d'un traitement à l'autre. Voici les limites de l'examen :

- l'analyse était limitée par des lacunes méthodologiques dans les études de séries de cas, comme le mauvais compte rendu des données;
- le recours à des interventions chirurgicales précises risque de limiter la généralisabilité des constatations;
- les constatations reposaient sur un faible nombre d'études.

Dans un rapport portant sur les données probantes concernant l'évaluation des essais sur échantillon non aléatoire et les méthodes employées à cette fin, Deeks et coll. (2003) ont

relevé 194 outils utilisés pour évaluer la qualité de ces essais. Soixante d'entre eux englobaient au moins cinq domaines de validité interne précisés au préalable sur six et ont été jugés supérieurs. Quatorze ont été classés comme étant les « meilleurs » et englobaient trois éléments centraux sur quatre qui revêtaient une importance particulière pour les études sur échantillon non aléatoire (répartition, groupes comparables, facteurs prévisionnels définis et recours à l'adaptation à la clientèle). Les auteurs ont établi que six des 14 outils convenaient aux examens systématiques. La force de ces outils était la formulation des éléments qui canalisait les réponses des examinateurs de manière systématique pour que les évaluations soient aussi objectives que possible.

Katrak et coll. (2004) ont réalisé l'examen systématique de 121 outils d'évaluation critique publiés à partir de 108 documents retrouvés grâce à l'interrogation de bases de données électroniques et d'Internet. La plupart des outils (87 %) étaient propres au modèle de recherche, et bon nombre d'entre eux (45 %) avaient été créés pour des études expérimentales (EEAC et essais cliniques sur échantillon contrôlé). Sur 16 outils génériques d'évaluation critique, six ont été mis au point pour des études expérimentales et par observation. Onze se sont avérés utiles pour tout modèle d'étude qualitative ou quantitative. Les auteurs ont extrait 74 éléments à partir de 19 outils d'évaluation critique pour des études par observation. Ces éléments étaient axés sur :

- les analyses de données,
- la prise en compte des variables confusionnelles,
- la taille de l'échantillon ou le calcul de la puissance,
- l'entreprise ou non d'une analyse statistique appropriée.

Ils ont extrait 36 éléments à partir de sept outils d'évaluation critique pour les études qualitatives. La plupart d'entre eux étaient axés sur l'évaluation de la validité externe, les méthodes d'analyse des données et la justification de l'étude. Ces outils ne renfermaient aucun élément sur la sélection des échantillons, la randomisation, le traitement à l'aveugle, l'intervention ou le biais. Les outils génériques, supposément utilisables dans le cas des études expérimentales ou par observation, contenaient des éléments axés sur l'échantillonnage et l'analyse des données, comme l'opportunité des analyses statistiques, la taille de l'échantillon et le calcul de la puissance. Cet examen n'a révélé aucun exemple idéal d'outil d'évaluation pour un type d'étude ou l'autre.

Résumé de la section

Ces examens systématiques ont intégré avec succès des études primaires où les données ne correspondaient pas à des groupes témoins, ou ils ne portaient pas sur des études avec outils d'évaluation critique pour l'évaluation de la qualité des études sans groupes témoins. Les examens touchant la possibilité de combiner les résultats d'EEAC et d'EENC portent à croire que les études bien conçues peuvent donner des résultats semblables, peu importe le genre d'étude. Bon nombre de ces examens portent à croire que c'est plutôt la mauvaise conception des études primaires, et non le modèle d'étude en soi, qui donne une note de qualité inférieure. Ils signalent que la sous-communication de la méthodologie, la mauvaise gestion des variables confusionnelles et le manque de terminologie analytique uniforme sont problématiques lorsqu'il s'agit de comparer les données d'une étude à l'autre.

Méthodes pour faire la synthèse d'études quantitatives sans groupes témoins

La présente section englobe cinq documents qui décrivaient des méthodes pour faire la synthèse d'études quantitatives sans groupes témoins.

Dans un document de 2005, Greenhalgh et coll. décrivaient leur processus d'examen méta-narratif, mis au point comme méthodologie pour faire la synthèse de données probantes d'une discipline à l'autre. Ils ont recouru à un processus en six étapes qui comprenait la planification, le dépouillement documentaire, la structuration, l'évaluation, la synthèse et la formulation de recommandations. Selon les auteurs, leur approche à la production de comptes rendus « graduels » des principales traditions de recherche a permis de comprendre la documentation complexe. Cinq grands principes sous-tendaient cette technique méta-narrative :

- le pragmatisme
- le pluralisme
- l'historicité
- la contestation
- l'examen par les pairs

Cette approche peut s'avérer utile lorsque la portée de l'étude est large et que la documentation est variée, et quand les chercheurs ont abordé un problème commun suivant divers modèles d'étude.

Whittemore et Knafl (2005) ont souligné que la méthode d'examen par intégration était une bonne solution pour intégrer des données provenant d'études menées suivant diverses méthodologies. Les auteurs ont exposé dans leurs grandes lignes des stratégies pour rehausser la rigueur des examens par intégration. Ils ont proposé cinq étapes pour mieux garantir la rigueur méthodologique de ce genre d'examen :

1. définition du problème : permet à la question d'étude d'être clairement définie;
2. dépouillement documentaire : intègre une stratégie de recherche exhaustive;
3. évaluation des données : devient un peu complexe en raison des études méthodologiquement diverses qui sont incluses dans l'examen; évaluer la qualité ne peut se faire suivant le processus standard d'examen systématique, mais il se peut qu'il faille se concentrer sur des exemples d'authenticité, la qualité méthodologique, la valeur informationnelle et la représentativité des études primaires disponibles;
4. analyse des données : comprend la réduction des données, la présentation, la comparaison et les conclusions; la réduction des données se fait par un système de classement global où les études primaires sont divisées en sous-groupes qui procurent un ordre logique pour l'analyse; ce classement en sous-groupes peut être fondé sur les genres de données probantes, la chronologie, le cadre, les caractéristiques de l'échantillon, ou peut se faire par un classement conceptuel prédéterminé; l'étape suivante du processus de réduction des données consiste à coder et à extraire les données dans un cadre réalisable; la présentation des données exige de convertir les données extraites au moyen de variables et de sous-groupes

précis; ces présentations englobent des matrices, des graphiques et des tableaux qui permettent de comparer les études entre elles; la comparaison des données permet de définir des modèles, des thèmes et des relations entre et parmi les variables; la dernière étape consiste à tirer les conclusions;

5. présentation : comprend l'exposition des constatations dans un examen par intégration qui peut être sous forme de tableaux ou de schémas et devrait inclure les conséquences sur la pratique, les politiques et la recherche.

Norris et Atkins (2005) ont étudié 49 examens publiés sur une période de cinq ans par les centres de pratique fondée sur des données probantes et qui incluaient des données probantes provenant d'études conçues autrement que comme des EEAC. Ces modèles englobaient des études de cohortes et des séries chronologiques, des études de séries de cas avant-après et des études transversales. Les auteurs ont laissé entendre qu'intégrer des études de cohortes et de séries de cas dans les examens offrait un avantage, soit celui que ces types de modèles sont plus à même de rendre les résultats à long terme que les EEAC. Ils ont observé que même si la plupart des 49 examens englobaient des études sur échantillon non aléatoire, les données probantes tendaient à provenir des EEAC. Évaluer la qualité d'essais sur échantillon non aléatoire constituait un défi. Sur les 49 examens, 25 % n'évaluaient pas la qualité, 16 % reposaient sur des listes de contrôle ou des barèmes de notation publiés, 10 % adaptaient des instruments de notation publiés et 49 % étaient fondés sur des listes de contrôle dressées par les examinateurs. Peu d'outils avaient fait l'objet d'un essai de validité. Afin de garantir le respect des principes des meilleures données probantes et de réduire le biais, Norris et Atkins ont suggéré que les examens commencent par un protocole détaillé qui serait alors respecté scrupuleusement du début à la fin. De plus, la stratégie de recherche pourrait devoir être plus fluide et reprise pour que le manque de stratégies de recherche adaptées et spécifiques dans les EEAC soit compensé. Les auteurs ont formulé les recommandations suivantes :

1. les examinateurs devraient évaluer la disponibilité des EEAC s'attachant à leur question d'examen avant de déterminer les critères d'inclusion finals;
2. lorsqu'ils envisagent d'inclure des essais sur échantillon non aléatoire, les examinateurs doivent songer aux biais possibles et se demander si ces biais peuvent être réduits au minimum dans des essais sur échantillon non aléatoire bien menés;
3. le processus d'examen doit être transparent afin que les motifs d'inclusion et d'exclusion des modèles d'étude soient clairs;
4. les examinateurs devraient songer à évaluer la qualité de chaque étude;
5. les examinateurs devraient prendre conscience de l'effet de leurs critères d'inclusion sur les constatations, et aborder les effets possibles du biais sur leurs conclusions;
6. la qualité méthodologique des études incluses doit faire partie de la discussion et des conclusions, peu importe le modèle d'étude.

Jackson et Waters (2005) ont examiné précisément les défis que représentait la rédaction d'examens systématiques pour le secteur de la santé publique. Les difficultés inhérentes comprenaient les interventions à composantes multiples, les mesures multiples de résultats, les populations variées et les mélanges de modèles d'étude. Ils ont convenu comme

d'autres qu'en santé publique, ce doit être l'intervention et la population qui dictent le modèle d'étude, et non l'inverse. Les chercheurs ont reconnu la controverse qui entoure l'opportunité du processus d'examen systématique dans le cas des études sur les interventions en santé publique. Ils ont aussi fait des suggestions pour améliorer le processus et en faire un moyen de modérer la critique. Chercher de la documentation est complexe et requiert beaucoup de temps. Par conséquent, il est essentiel de prévoir assez de temps pour une recherche adéquate, et d'interroger de nombreuses bases de données électroniques et sources de littérature grise. Il faut absolument évaluer la qualité de tous les articles inclus, et ils ont recommandé d'utiliser l'outil mis au point par le Projet de pratique efficace en santé publique, à Hamilton (Ontario) (Thomas et coll., 2004a), soit le « Quality Assessment Tool for Quantitative Studies ». Ce solide instrument (Deeks et coll., 2003) peut servir à des EEAC et des études quasi-expérimentales sur échantillon non contrôlé. Porter attention aux cadres théoriques dans les études primaires et les examens systématiques permet d'expliquer les différences entre la planification et les résultats des interventions. Mesurer l'intégrité des interventions peut aider à déterminer si une intervention a été inefficace en raison d'une mauvaise planification ou parce que la prestation de l'intervention était incomplète. Étant donné les diverses populations qui font l'objet d'interventions en santé publique, les chercheurs devraient s'attendre à de l'hétérogénéité. Les chercheurs et les examinateurs ne trouveront pas déjà tous ces éléments dans toutes les études primaires, mais en prendre conscience pourra aider à renforcer les examens systématiques découlant de la recherche sur les interventions en santé publique.

Le document de 1998 d'Atkins et DiGuseppi portait sur la nécessité d'effectuer des recherches sur l'utilisation des services de santé préventifs. Il n'était pas toujours possible d'obtenir des données probantes issues d'EEAC dans le domaine de la promotion de la santé et des services de santé préventifs. Les chercheurs ont mentionné les problèmes connus qui sont associés aux études par observation comme motif de prudence. Cependant, ils ont reconnu que ces études peuvent augmenter les connaissances. Les forces des études par observation comprennent :

- l'établissement de liens dans l'enchaînement de causalité;
- la compréhension de l'histoire naturelle de la maladie;
- la définition des facteurs de risque;
- la mesure du respect des traitements et des effets négatifs de ceux-ci;
- l'établissement de l'exactitude des tests de diagnostic;
- l'évaluation de l'efficacité des interventions.

Le risque de biais peut être modifié grâce à des échantillons représentatifs fondés sur la population et à l'idée de bien tenir compte des variables confusionnelles. Les chercheurs doivent aussi reconnaître que les études par observation tendent à exagérer les effets bénéfiques du traitement. Les auteurs recommandent d'inclure un grand nombre d'études bien conçues où les effets sont constants et importants de préférence quand il s'agit d'utiliser des études par observation pour formuler des recommandations de traitement.

Résumé de la section

Il est difficile de faire la synthèse de données quantitatives issues d'études sans groupes

témoins et de les intégrer. Les six documents faisaient état de diverses méthodes pour intégrer ce genre de données de manière utile. Les auteurs ont convenu qu'il est avantageux d'intégrer des données probantes provenant de divers modèles d'étude, mais il est primordial d'évaluer la qualité afin d'obtenir des données probantes dignes de confiance. Il serait peut-être utile d'intégrer dans les examens systématiques les études primaires qui définissent clairement les biais possibles et les effets probables.

Combiner des études qualitatives et quantitatives

La présente section vise à examiner trois articles qui décrivent des méthodes pour faire la synthèse des données provenant d'études qualitatives et quantitatives.

Déterminer la qualité d'études non expérimentales est difficile et controversé. Cependant, ces genres d'études renferment souvent des renseignements importants pour la santé publique. Évaluer la qualité à l'aide d'outils standard n'est pas particulièrement utile quand peu de ces études obtiendraient une note élevée ou moyenne. Harden et coll. (2004) ont utilisé une approche standard à l'évaluation de la qualité méthodologique dans des examens systématiques portant sur 35 études primaires sur les non-interventions. Seulement quatre d'entre elles ont répondu à tous les critères utilisés pour évaluer la qualité. Ils ont mis au point un outil de récupération de données pour mieux déconstruire chaque étude (quantitative ou qualitative). De là, ils ont reconstruit les résultats et les ont présentés sous forme de résumés structurés et de tableaux de données probantes. Le processus de synthèse était non linéaire et faisait intervenir deux examinateurs faisant l'aller-retour entre les documents, l'extraction des données et les tableaux de données probantes. Il n'était pas approprié de regrouper les résultats, comme dans une méta-analyse. Ces chercheurs ont plutôt employé des méthodes combinées suivant des thèmes définis pour faire la synthèse des constatations. Afin de répondre à la principale question de recherche, les examinateurs ont intégré les constatations issues des études sur les non-interventions dans 36 études sur les interventions, et comparé les résultats afin d'établir les ressemblances et les différences (Oliver et coll., 2005). Recourir à ces deux genres de résultats a permis un plus grand éventail de points de vue et de mieux comprendre les enjeux de santé publique du point de vue des gens qui font l'objet des interventions ciblées.

Les méta-examens, comme les décrivent Wong et Raabe (1996), intègrent des données qualitatives et quantitatives. Les auteurs ont suggéré de recourir à un méta-examen pour s'attaquer à certaines des limites associées aux examens qualitatifs traditionnels, comme la subjectivité de la sélection des articles et le poids attribué à chaque étude ainsi que l'absence d'analyse quantitative. On y est arrivé en incluant un examen narratif dans la méta-analyse des données quantitatives. Le premier facteur pour déterminer si une étude devrait être incluse est le modèle de l'étude; les études de modèle différent ne devraient pas être groupées aux fins de méta-analyse. Celle-ci peut augmenter la puissance statistique, mais elle ne peut compenser les questions méthodologiques reliées aux modèles d'étude. Il y a des étapes de base à suivre pour effectuer un méta-examen :

1. définir la question de recherche;
2. procéder à un examen documentaire exhaustif;
3. créer des critères d'inclusion a priori;

4. effectuer un examen qualitatif traditionnel;
5. effectuer une méta-analyse quantitative;
6. intégrer un examen qualitatif traditionnel avec une méta-analyse quantitative;
7. appliquer les critères de causalité à l'interprétation.

Les auteurs laissent entendre qu'intégrer un examen narratif traditionnel dans une méta-analyse peut réduire la subjectivité.

Un document de 2007 publié par Sandelowski, Barroso et Voils portait sur la nécessité pour les disciplines de la santé d'inclure des constatations d'études qualitatives et quantitatives dans les examens, surtout en raison du nombre croissant d'études fondées sur les méthodes mélangées. Les auteurs ont examiné 42 rapports, y compris des articles de revue, des dissertations ou des thèses non publiées et des rapports techniques. Afin de différencier leur approche d'un examen systématique standard, aucun rapport n'a été exclu d'après la qualité méthodologique. Pour faire la synthèse de leurs constatations, ils ont mis au point un modèle de méta-synthèse qualitative où les constatations provenant d'études primaires sont groupées par sujet et par thème. Le méta-résumé comprenait l'extraction, le groupage et le formatage des constatations, ainsi que le calcul de l'ampleur de la fréquence et de l'effet pour les données quantitatives. Cette approche peut servir à faire la synthèse des enquêtes fondées sur des méthodes mélangées où la collecte des données et les procédures d'analyse sont semblables.

Résumé de la section

Intégrer des données qualitatives et quantitatives dans un rapport peut s'avérer utile lorsqu'il s'agit de comprendre le succès ou l'échec d'interventions. Toutefois, évaluer la qualité d'études sans groupes témoins peut être problématique, surtout si la qualité est évaluée par des méthodes traditionnelles employées dans les examens systématiques ou les méta-analyses. Les résultats de l'évaluation de la qualité n'ont pas servi de critères d'exclusion. Ils ont plutôt été intégrés dans le débat sur les constatations. Grouper les constatations selon une méthode thématique et globale est une autre approche appropriée à la méta-analyse. Les chercheurs pourraient intégrer certains processus d'examen systématique, comme demander à deux personnes d'examiner et d'extraire les données, mais la plupart des auteurs suggèrent d'utiliser une approche narrative à la présentation des résultats lorsque la méta-analyse ne convient pas.

Études comparant les résultats selon des méthodes différentes

La recherche comprenait deux études qui visaient à examiner les problèmes qui se présentent quand il s'agit de comparer les résultats d'études sur échantillon aléatoire ou non aléatoire concernant la même intervention.

Jefferson et Demicheli (1999) ont évalué la capacité de différents modèles d'étude à vérifier quatre aspects du rendement des vaccins (l'immunogénicité, la durée de l'immunité apportée, l'incidence et la gravité des effets secondaires et le nombre d'infections prévenues par la vaccination). Leurs constatations révèlent qu'en vaccinologie, les modèles d'étude expérimentaux et non expérimentaux sont souvent complémentaires. Cependant, dans

certains cas, la qualité du vaccin pourrait seulement être mesurée par un genre d'étude. Par exemple, une EEAC est préférable pour mesurer les effets secondaires. Les modèles non expérimentaux conviennent quand :

- une expérience est impossible, inutile ou inappropriée;
- l'efficacité individuelle doit être mesurée d'après des événements négatifs et peu fréquents;
- les interventions préviennent de rares cas ou l'efficacité d'une intervention au sein d'une population doit être mesurée à long terme.

En recourant à des données d'entrevue recueillies auprès de 17 auteurs et utilisateurs d'ÉTS, Rotstein et Laupacis (2004) ont jeté de la lumière sur les différences entre les ÉTS et les examens systématiques. Voici quelles sont ces différences :

1. les normes méthodologiques — les ÉTS peuvent inclure de la documentation de mauvaise qualité méthodologique si un sujet est important pour les décideurs;
2. la reprise d'études antérieures — les examens systématiques n'ont pas à être repris si les études antérieures étaient de grande qualité ou quand il n'existe aucune nouvelle donnée probante de grande qualité; dans le cas des ÉTS, il faut souvent reprendre des études pour défendre les conclusions du rapport;
3. le choix des sujets — les sujets sont plus axés sur les politiques avec les ÉTS, alors que les examens systématiques tendent à être orientés par des questions d'efficacité;
4. l'inclusion de spécialistes du contenu (dans les examens systématiques) et de responsables de l'élaboration des politiques (dans les ÉTS) comme auteurs;
5. l'inclusion d'évaluations économiques, dans les ÉTS;
6. la formulation de recommandations d'orientation, dans les ÉTS;
7. la diffusion du rapport, réalisée plus souvent activement dans le cas des ÉTS.

Les ÉTS ne sont pas précisément conçues pour évaluer des données probantes scientifiques, alors que les examens systématiques sont destinés à cette fin. Cependant, la nature impartiale de l'enquête scientifique aiderait à renforcer les décisions stratégiques découlant de données probantes issues d'ÉTS. Les niveaux différents de données probantes (sous les EEAC dans la hiérarchie) sont plus acceptables dans les ÉTS.

Résumé de la section

Ces deux comparaisons de résultats de la même intervention obtenus par des modèles différents font ressortir qu'il est utile d'envisager des modèles différents dans les examens, non seulement parce qu'il pourrait s'agir des meilleurs renseignements qui soient, mais aussi parce qu'ils sont plus susceptibles de répondre aux questions relatives à l'efficacité à long terme, aux événements négatifs et aux cas rares. Les forces que représente le fait d'inclure d'autres modèles l'emportent sur les limites.

Discussion

Le principal point de mire du présent document sur le plan de la recherche était les méthodes d'inclusion des résultats provenant d'essais sur échantillon non aléatoire, ainsi que la qualité et la synthèse de ces données. Il s'agit d'une importante question de recherche, car dans bien des domaines de la santé et de la santé publique, les données probantes provenant d'EEAC n'existent pas. Les données probantes issues d'essais sur échantillon non aléatoire ont toujours été critiquées comme étant peu fiables, en raison du risque de biais accru. Il est couramment entendu que le biais de sélection, de répartition ou d'attrition a tendance à être plus grand dans les essais sur échantillon non aléatoire. Les lecteurs de données issues d'essais sur échantillon non aléatoire doivent être conscients que ces biais peuvent exister, et les chercheurs devraient tenir compte de biais possibles dans leurs études. Voici quelques suggestions découlant des études examinées dans le présent document :

- dans tous les essais sur échantillon non aléatoire, un modèle d'étude solide et transparent s'impose;
- la question de recherche devrait déterminer le modèle d'étude;
- les auteurs d'étude doivent intégrer des renseignements détaillés sur leur étude; un manque d'information signifie que les lecteurs ne peuvent déterminer ce qui a été fait ou non dans le modèle d'étude;
- il faut que la terminologie et les descriptions soient uniformes d'une étude sur échantillon non aléatoire à l'autre pour que les comparaisons se fassent facilement.

Il existe plusieurs approches méthodologiques à l'inclusion de données provenant d'études sur échantillon non aléatoire. Nous avons découvert que c'est le travail de Whitemore et Knafl (2005) qui est le plus systématique. Whitemore et Knafl proposent une méthode d'examen par intégration en cinq étapes qui intègre la définition du problème, le dépouillement documentaire ainsi que l'évaluation, l'analyse et la présentation des données.

Bon nombre de chercheurs utilisent divers outils d'évaluation pour déterminer la qualité des études incluses. Norris et Atkins (2005) fournissent une liste exhaustive de recommandations pour évaluer la qualité d'étude qui est ressortie des 49 examens sur lesquels ils se sont penchés. Ces recommandations étaient exposées dans leurs grandes lignes dans le présent document. D'autres auteurs ont utilisé des listes de contrôle ou des instruments de notation existants, comme MINORS ou GRADING. Certains ont adapté des outils pour qu'ils soient conformes aux modèles d'étude découverts dans les études primaires, alors que d'autres ont créé de nouveaux outils qui correspondaient aux renseignements à extraire des études incluses. La validité ou la fiabilité de bon nombre de ces outils n'a pas été éprouvée. Nous recommandons que d'autres études soient effectuées pour que soient établies la validité et la fiabilité de ces outils d'évaluation de la qualité dans le cas des essais sur échantillon non aléatoire. Autant que possible, les examinateurs devraient recourir à des outils éprouvés et jugés valides et fiables en ce qui touche les essais sur échantillon non aléatoire, comme le Quality Assessment Instrument for Primary Studies (outil d'évaluation de la qualité pour les études primaires) mis au point par le Projet de pratique efficace en santé publique, à Hamilton (Ontario).

Inclure des données provenant d'essais sur échantillon non aléatoire s'avère difficile. Cependant, leur inclusion peut apporter plusieurs avantages. S'inspirer d'études sur échantillon non aléatoire signifie que les chercheurs peuvent ajouter un éventail plus large de documentation à leurs stratégies de recherche (Thomas et coll., 2004). Intégrer ces études peut élargir la portée des questions auxquelles il est possible de répondre. Cette recherche permet d'établir pourquoi, comment et pour qui les interventions ou les programmes sont fructueux ou non (Oliver et coll., 2005).

Recommandations

Actuellement, il n'existe aucun modèle standard pour faire la synthèse des études sans groupes témoins. Le présent document a pour objet d'examiner les méthodes de synthèse, les outils d'évaluation critique et les études portant sur l'évaluation et la synthèse des études sans groupes témoins.

Le présent document exposait dans leurs grandes lignes certaines précautions à prendre quand il s'agit d'utiliser des études sur échantillon non contrôlé comme données probantes, et il recommandait des outils et des processus.

1. Intégrer des études sans groupes témoins dans les examens systématiques, en particulier quand il n'y a aucune autre étude à prendre en compte. Les études sans groupes témoins peuvent aussi renseigner sur l'efficacité à long terme, les cas rares et les effets néfastes.
2. Éviter de généraliser l'orientation des écarts de revenu entre les études sur échantillon aléatoire et celles sur échantillon non contrôlé. Bien que les effets soient plus souvent de moindre envergure dans les essais sur échantillon aléatoire et contrôlé (EEAC), certains donnent des résultats inverses ou identiques quelle que soit le modèle d'étude.
3. Adopter l'approche de Whitemore et Knafl (2005) selon laquelle il faut inclure les résultats d'essais sur échantillon non aléatoire. Whitemore et Knafl proposent une méthode d'examen par intégration en cinq étapes qui englobe la définition des problèmes, le dépouillement documentaire ainsi que l'évaluation, l'analyse et la présentation des données.
4. Recommandations concernant les méthodes d'examen systématique, incluant les études sans groupes témoins :
 - a) utiliser un outil d'évaluation critique dont l'uniformité interne, la fiabilité de test-retest, le coefficient d'objectivité et, à tout le moins, la validité apparente et de critère ont été éprouvés (Downs et Black, 1998; Slim, Nini, Forestier et coll., 2003; Steuten, Vrijhoef, van-Merode et coll., 2004; Thomas, Ciliska, Dobbins et coll., 2004a; Zaza, Wright-De Agüero, Briss et coll., 2000);
 - b) éviter de méta-analyser les résultats d'études par observation.
5. Suggestions aux auteurs d'études primaires :
 - a) assurer un modèle d'étude solide et transparent dans tous les essais sur échantillon non aléatoire;
 - b) confirmer que la question de recherche détermine le modèle d'étude;
 - c) inclure des renseignements détaillés sur l'étude; s'il manque de l'information, les lecteurs ne pourront déterminer ce qui a été fait ou non dans le modèle d'étude;
 - d) garantir l'uniformité de la terminologie et des descriptions dans toutes les études sur échantillon non aléatoire afin de faciliter les comparaisons.
6. Répercussions sur les recherches ultérieures :

- a) effectuer des essais de fiabilité et de validité sur les outils d'évaluation critique;
 - b) mener d'autres études pour évaluer les divers résultats obtenus par des essais sur échantillon aléatoire ou non.
7. Il est difficile d'inclure des données provenant d'essais sur échantillon non aléatoire; cependant, cela peut offrir plusieurs avantages :
- a) un ensemble de documentation plus large peut être inclus dans les stratégies de recherche (Thomas, Harden, Oakley et coll., 2004);
 - b) la portée des questions auxquelles on peut répondre peut être élargie;
 - c) l'étude permet d'établir pourquoi, comment et pour qui les interventions ou les programmes réussissent ou échouent (Oliver, Harden, Rees et coll., 2005).

Tableau résumant les résultats des documents pertinents

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Rangel et coll. 2003 États-Unis OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Décrire la mise au point et les applications possibles d'une échelle normalisée d'évaluation de la qualité destinée aux études rétrospectives en chirurgie pédiatrique.</p>	<p>Ils ont mis au point un instrument d'évaluation exhaustive de la qualité intégrant 30 éléments à l'intérieur de trois sous-échelles.</p> <p>Les sous-échelles ont été conçues pour évaluer la pertinence clinique, la méthodologie de communication et la force des conclusions.</p> <p>Les cotes de qualité globale (mauvaise, assez bonne ou bonne) ont été dérivées de la combinaison des notes obtenues dans chaque sous-échelle.</p> <p>Six examinateurs indépendants ont examiné le coefficient d'objectivité en évaluant l'instrument dans 10 études rétrospectives tirées de documentation sur la chirurgie pédiatrique.</p> <p>Constatations</p> <ul style="list-style-type: none"> • L'instrument avait un excellent coefficient d'objectivité. • Les degrés d'entente étaient très élevés pour tous les éléments de l'instrument (84,6 % de concordance, n = 1573 éléments) et toutes les sous-échelles (échelle de 73,3 % à 85,8 %, n = 60 à 1258) 	<p>Conclusions</p> <ul style="list-style-type: none"> • Les auteurs ont mis au point une échelle d'évaluation de la qualité normalisée et fiable pour l'analyse de données rétrospectives en chirurgie pédiatrique. • Les résultats peuvent ne pas être applicables à d'autres interventions. <p>Applications possibles</p> <ul style="list-style-type: none"> • Fournir aux chirurgiens praticiens une base de connaissances pour évaluer de manière critique les données rétrospectives publiées. • Procurer une méthodologie normalisée pour l'examen systématique des données rétrospectives existantes. • Élaborer des lignes directrices de communication normalisées pour les revues examinées par les pairs.
<p>Margetts et coll. 1995 Royaume-Uni OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Créer un système de notation pour mieux juger de la qualité scientifique des études épidémiologiques par observation reliant le régime alimentaire au risque de cancer.</p>	<p>Le système de notation a été créé d'après des rubriques clés utilisées pour élaborer des protocoles de recherche et comprenait des questions sous les rubriques :</p> <ul style="list-style-type: none"> • trois pour les études cas-témoin (évaluation de l'alimentation, recrutement de sujets et analyse) – système de notation à l'annexe A; • quatre pour les études de cohortes (évaluation de l'alimentation, définition de cohorte, vérification et analyse) – système de notation à l'annexe B. <p>La variation d'un observateur à l'autre a été évaluée.</p> <p>Les observateurs s'entendaient bien sur le classement des études.</p> <p>Le système de notation a été appliqué à un examen du rapport entre la viande et le risque de cancer :</p> <p>D'après ce qui semblait un vaste échantillon de documentation, relativement peu d'études ont obtenu une bonne note (supérieure à 65 %). Cependant, ces études tendaient à fournir des renseignements plus uniformes.</p> <p>Dans le cas des études cas-témoin, 34 des 106 études ont obtenu une note supérieure à 65 %.</p> <p>Dans celui des études de cohortes, 10 des 41 études ont obtenu une note supérieure à 65 %.</p>	<p>Le prototype de système de notation dont il est question ici a permis aux auteurs de décrire les données épidémiologiques sur le régime alimentaire et le cancer, mais il n'est pas applicable à d'autres cadres ou interventions</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Downs et Black 1998 Royaume-Uni</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Vérifier la faisabilité de créer une liste de contrôle valide et fiable ayant les caractéristiques suivantes :</p> <p>qui convient à l'évaluation des études sur échantillon aléatoire ou non aléatoire; qui donne à la fois une note globale pour la qualité de l'étude et un profil de notes, non seulement pour la qualité de la communication, la validité interne et la puissance, mais aussi pour la validité externe.</p>	<p>L'étude a commencé par une version pilote de l'outil. Les chercheurs ont demandé à trois experts d'évaluer la validité apparente et celle du contenu. Ensuite, deux évaluateurs ont utilisé l'outil pour évaluer 10 études sur échantillon aléatoire et 10 études sur échantillon non aléatoire afin de mesurer la fiabilité.</p> <p>Constatations</p> <ul style="list-style-type: none"> • À l'aide d'évaluateurs différents, on a révisé la liste de contrôle et vérifié sa cohérence interne (Kuder-Richardson 20), sa fiabilité de test-retest et son coefficient d'objectivité, la validité des critères et le fardeau du répondant. • Les chercheurs ont constaté une forte cohérence interne du point de vue de l'indice de qualité. • La fiabilité de test-retest et le coefficient d'objectivité de l'indice de qualité étaient bons. 	<p>Conclusions</p> <ul style="list-style-type: none"> • Il est possible de produire une liste de contrôle qui évalue les études sur échantillon aléatoire ou non aléatoire. • La principale préoccupation concerne la validité externe de la liste et son application au-delà de l'objet d'étude.
<p>Slim et coll. 2003 France</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Mettre au point et valider un indice méthodologique pour les études sur échantillon non aléatoire (MINORS) que les lecteurs, les examinateurs de manuscrits ou les rédacteurs en chef de revue pourraient utiliser pour évaluer la qualité de ce genre d'études.</p>	<p>L'indice comptait 12 éléments.</p> <p>Dans le cas des études non comparatives, voici quels étaient les éléments :</p> <ul style="list-style-type: none"> • un objectif clairement énoncé, • l'inclusion de patients consécutifs, • une collecte prospective des données, • des critères convenant à l'objectif de l'étude, • une évaluation impartiale des critères de l'étude, • une période de suivi appropriée à l'objectif de l'étude, • un suivi impossible dans moins de 5 % des cas. <p>Dans le cas des études comparatives, voici quels étaient les éléments :</p> <ul style="list-style-type: none"> • un groupe témoin ayant fait l'objet d'un exemple idéal d'intervention, • des groupes contemporains, • une équivalence de base des groupes, • un calcul prospectif de la taille de l'étude, • des analyses statistiques adaptées au modèle d'étude. <p>Les examinateurs s'entendaient bien sur l'indice: la fiabilité de test-retest était élevée, d'après le coefficient Kappa; et la cohérence interne était bonne, selon le coefficient alpha de Cronbach.</p>	<p>MINORS est un instrument valide destiné à évaluer la qualité méthodologique des études d'interventions chirurgicales sur échantillon non aléatoire, qu'elles soient comparatives ou non.</p> <p>Il se pourrait que les chercheurs doivent faire preuve de prudence en appliquant l'instrument à des interventions non chirurgicales.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Zaza et coll. 2000 États-Unis</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p>	<p>1. Contribuer à la cohérence, réduire le biais et améliorer la validité pour les chercheurs qui utilisent le Guide to Community Prevention Services: Systematic Reviews and Evidence-Based Recommendations (guide des services communautaires de prévention : examens systématiques et recommandations fondées sur des données probantes) (le guide).</p> <p>2. Élaborer un formulaire normalisé d'extraction des données à utiliser avec le guide.</p>	<p>Le formulaire a été élaboré par :</p> <ul style="list-style-type: none"> examen des méthodologies employées dans d'autres examens systématiques; communication des normes établies par les revues de santé et de sciences sociales; examen de la documentation sur les évaluations, les statistiques et les méta-analyses; sollicitation d'opinions d'experts. <p>Le formulaire a servi à évaluer la qualité méthodologique d'études primaires (23 questions) et à classer et décrire les principales caractéristiques de l'intervention et de l'évaluation (26 questions).</p> <p>Les chercheurs ont examiné des procédures et des résultats d'étude; et ils ont évalué les menaces à la validité dans six catégories :</p> <ul style="list-style-type: none"> descriptions d'intervention et d'étude échantillonnage mesure analyse interprétation des résultats autres questions d'exécution 	<p>Le processus a permis d'améliorer la validité et la fiabilité du guide.</p> <p>Le formulaire normalisé d'extraction de données peut aider les chercheurs et les lecteurs d'études primaires à examiner le contenu et la qualité des études. Il peut aussi aider à élaborer des manuscrits à soumettre à des revues examinées par les pairs.</p>
<p>Greer et coll. 2000 US</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p> <p>(Systèmes de notation des données probantes)</p>	<p>Décrire en détail le système de notation des données probantes et des conclusions qui a été mis au point par l'Institute for Clinical Systems Improvement (ICSI) à l'intention des cliniciens praticiens qui rédigent les documents et s'en servent pour prendre des décisions sur les soins à prodiguer aux patients.</p>	<p>Le système de notation comprenait une évaluation de chaque compte rendu de recherche et une évaluation de la force globale des données probantes appuyant une conclusion ou une recommandation en particulier.</p> <p>Le système de notation servant à évaluer la qualité des comptes rendus de recherche se comprend facilement et est plus facile à utiliser pour les médecins praticiens.</p> <p>Le système de notation a pour objectif de faire ressortir un compte rendu qui est anormalement fort ou remarquablement déficient.</p> <p>Le système de notation des données probantes :</p> <ul style="list-style-type: none"> comprend le tableau de notation des conclusions, tableau qui exige l'énoncé d'une conclusion, un résumé des comptes rendus de recherche qui sont à l'appui de la conclusion ou la réfutent, l'attribution de catégories et d'indices de qualité aux comptes rendus de recherche, et l'attribution d'une note à la conclusion; a servi à rédiger plus de 40 lignes directrices et de nombreux rapports d'évaluation technologique. 	<p>Conclusions</p> <ul style="list-style-type: none"> Le système semble permettre de surmonter la complexité de certains systèmes publiés de notation des données probantes, bien qu'il donne toujours un classement défendable des conclusions fondées sur la force des données probantes sous-jacentes. La fiabilité du système doit être rigoureusement mise à l'épreuve.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
Steuten et coll. 2004 Pays-Bas OUTILS D'ÉVALUATION CRITIQUE	Décrire dans quelle mesure les instruments actuels sont utiles lorsqu'il s'agit d'évaluer la qualité méthodologique de l'évaluation de l'évaluation des technologies de la santé (ÉTS) pour la gestion des maladies.	Méthodes ou résultats Des problèmes ont été relevés dans l'évaluation de la qualité méthodologique de six ÉTS pour la gestion des maladies dans le cas de trois instruments différents. Ils concernaient surtout : <ul style="list-style-type: none"> • le modèle d'étude (EEAC ou études par observation); • les critères de sélection et la limitation des patients; • les mesures de base et de résultats (plusieurs paramètres s'imposent); • le traitement à l'aveugle des patients et des fournisseurs (ce qui est impossible en gestion des maladies); • la description des interventions complexes à facettes multiples et l'évitement des co-interventions. Les chercheurs ont proposé et validé un nouvel instrument, HTA-DM (ÉTS-GM), qui : <ul style="list-style-type: none"> • compte quatre composantes (population à l'étude, description de l'intervention, mesure des résultats et analyse et présentation des données); • renferme 15 éléments (sur la validité interne, la validité externe et les considérations statistiques). 	Commentaires ou enjeux
Chou et Helfand 2005 États-Unis MÉTHODES (Examens systématiques (ES) sur les préjudices) OUTILS D'ÉVALUATION CRITIQUE	Examiner les défis méthodologiques de l'exécution d'ES sur les préjudices et souligner des exemples d'approches à leur égard provenant de 96 comptes rendus de données probantes sur les centres de pratique fondée sur des données probantes.	Certaines études ont permis de constater que les études par observation et les EEAC indiquent des estimations d'effets semblables. D'autres ont permis de constater que les essais cliniques indiquent des risques accrus d'événements négatifs que les études par observation (peut-être en raison de l'évaluation des préjudices qui est plus mauvaise dans les études par observation, ou parce qu'elles sont plus susceptibles d'être publiées si elles indiquent de bonnes méthodes ou de bons résultats). Inclure de grandes bases de données pourrait fournir des renseignements utiles sur les préjudices. Les réseaux fondés sur la pratique fournissent souvent des détails plus riches que les bases de données administratives — il est possible de définir et de mesurer les variables confusionnelles probables avec plus de confiance, mais elles sont plus difficiles à trouver par dépeuplement électronique parce que bon nombre des analyses du genre sont à propriété exclusive. Inclure des rapports de cas peut aider à définir les effets indésirables peu courants, inattendus ou à long terme des médicaments qui sont souvent différents de ceux décelés pendant les essais cliniques. Évaluer la qualité des rapports de préjudices : <ul style="list-style-type: none"> • Plusieurs ES ont permis de découvrir que le fait d'opter pour des études cas-témoin prospectives ou rétrospectives ou des études de cohortes, ou encore des études de séries de cas de petite 	Il faut disposer de meilleures données sur les préjudices pour effectuer des ES équilibrés. Il y a lieu de mener d'autres recherches afin d'établir de manière empirique l'effet d'inclure des données issues de diverses sources sur les évaluations de préjudices, mais aussi d'améliorer et de vérifier les critères de notation de la qualité des rapports de préjudices. Il importe : <ul style="list-style-type: none"> • d'éviter les combinaisons statistiques de données qui sont inappropriées; • de décrire avec soin les caractéristiques et la qualité des études incluses; • d'examiner à fond les sources possibles d'hétérogénéité.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>McIntosh et coll. 2004 Royaume-Uni MÉTHODES (ES)</p>	<p>Présenter l'expérience qui consiste à effectuer des examens systématiques (ES) sur les effets nocifs et à faire des suggestions de pratique et d'études pour l'avenir.</p>	<p>Les auteurs ont évalué les méthodes employées dans trois ES, en se concentrant sur la question d'examen, les modèles d'étude et l'évaluation de la qualité.</p> <p>Question d'examen</p> <ul style="list-style-type: none"> • Dans un examen, on posait une question précise axée sur la prestation de renseignements sur des effets nocifs spécifiques dans le but de produire un modèle économique; dans les deux autres examens, les questions étaient plus larges. <p>Modèles d'étude</p> <ul style="list-style-type: none"> • Les trois incluaient des données d'études sur échantillon aléatoire et par observation, mais chacun définissait les critères d'inclusion différemment. <p>Évaluation de la qualité :</p> <ul style="list-style-type: none"> • listes de contrôle publiées et appliquées — problèmes survenus; • communication inadéquate des caractéristiques fondamentales de conception des études primaires — listes de contrôle omettant des caractéristiques clés, comme la manière dont les effets nocifs étaient enregistrés. 	<p>Principaux domaines méritant des améliorations :</p> <ul style="list-style-type: none"> • concentrer la question d'examen sur des points précis, car les questions larges et imprécises n'aident pas au processus d'examen; • élaborer des méthodes normalisées pour évaluer la qualité des études sur les effets nocifs.
<p>Atkins et coll. 2005 États-Unis OUTILS D'ÉVALUATION CRITIQUE (Système de notation)</p>	<p>Mener un essai pilote et améliorer l'approche GRADE à la notation des données probantes et des recommandations.</p>	<p>L'approche GRADE a été créée par le GRADE Working Group pour évaluer la qualité des données probantes et la force des recommandations.</p> <p>Les chercheurs ont utilisé 12 profils de données probantes dans cette étude pilote.</p> <p>Chaque profil a été créé d'après les renseignements fournis par un examen systématique.</p> <p>Dix-sept personnes ont évalué indépendamment la qualité des données probantes et la force des recommandations pour chacun des 12 profils.</p> <p>Qualité des données pour chaque résultat</p> <ul style="list-style-type: none"> • Les chercheurs ont découvert qu'en plus du modèle d'étude, de la qualité, de la cohérence et de la franchise, d'autres critères de qualité ont aussi influencé le jugement sur les données probantes; les données éparées, les solides associations, les biais de publication, la relation dose-effet et les situations où toutes les variables confusionnelles plausibles ont renforcé plutôt qu'affaibli la confiance à l'égard de l'orientation de l'effet. <p>Fiabilité</p> <ul style="list-style-type: none"> • Sur la qualité des données probantes pour les résultats touchant chacune des douze questions (le coefficient de Kappa pour la concordance autre que fortuite allait de 0 à 0,82), on s'entendait dans diverses mesures. • Sur l'importance relative de chaque résultat, on s'entendait assez bien. • Sur l'équilibre entre les avantages, les préjudices et les recommandations, on s'entendait peu (ce qui pourrait s'expliquer en partie par l'accumulation de tous les écarts antérieurs dans la notation de la qualité et de l'importance des données probantes). • La plupart des écarts ont été facilement résolus par la discussion. 	

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Khan et coll. 2001 Royaume-Uni OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Fournir des renseignements généraux sur la qualité des études et décrire comment dresser des listes de contrôle pour l'évaluation de la qualité.</p>	<p>Les chercheurs ont décrit des raisons d'inclure l'évaluation de la qualité des études dans les examens, les types de biais, certaines méthodes pour se protéger contre le biais dans les études d'efficacité primaires et comment les instruments d'évaluation de la qualité devraient être mis au point.</p> <p>Les critères de qualité pour l'évaluation des études par observation comprennent le fait de répondre à une série de questions sur les études de cohortes, les études cas-témoins et les séries de cas.</p> <p>Études de cohortes</p> <ul style="list-style-type: none"> • Les groupes et la répartition des facteurs prévisionnels sont-ils suffisamment décrits? • Les groupes sont-ils réunis en un même endroit pendant la progression de la maladie? • L'intervention ou le traitement est-il déterminé avec fiabilité? • Les groupes étaient-ils comparables du point de vue de tous les facteurs confusionnels importants? • Les effets de ces variables confusionnelles ont-ils été adaptés adéquatement? • Y avait-il une relation dose-effet entre l'intervention et le résultat démontré? • Le résultat a-t-il été évalué sans que l'état d'exposition soit pris en compte? • Le suivi a-t-il été suffisamment long pour que les résultats soient obtenus? • La proportion de la cohorte a-t-elle fait l'objet d'un suivi? • Les taux et les motifs d'abandon étaient-ils semblables d'un groupe visé par l'intervention et d'un groupe non exposé à l'autre? <p>Études cas-témoins</p> <ul style="list-style-type: none"> • La définition des cas est-elle claire? • Le problème de santé des cas a-t-il été évalué et validé avec fiabilité? • Les témoins ont été choisis au hasard à même les cas? • En quoi les cas et les témoins sont-ils comparables en ce qui touche les facteurs confusionnels possibles? • Les interventions et les autres expositions ont-elles été évaluées de la même manière autant pour les cas que pour les témoins? 	<p>Conclusions</p> <ul style="list-style-type: none"> • Dans un examen où ces modèles d'étude sont intégrés, il peut être difficile de déterminer si les différences découlent de l'intervention ou de l'incompatibilité des groupes. • Les résultats devraient être considérés avec prudence.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Thomas et coll. 2004 Canada OUTILS D'ÉVALUATION CRITIQUE MÉTHODES (ES)</p>	<p>Décrire quatre enjeux relatifs aux examens documentaires systématiques sur l'efficacité des interventions en matière de services infirmiers de santé publique :</p> <ol style="list-style-type: none"> 1. processus d'examen systématique de la documentation; 2. mise au point d'un instrument d'évaluation de la qualité; 3. les méthodes et les résultats du PPESP jusqu'à présent; 4. quelques méthodes et résultats de la diffusion employée. 	<p>Domaines d'évaluation des études par observation</p> <ul style="list-style-type: none"> • Question d'étude • Population à l'étude • Comparabilité des sujets • Exposition ou intervention • Mesure à l'aveugle des résultats • Analyse statistique • Résultats • Discussion • Fondement <p>Composants de l'outil d'évaluation de la qualité du PPESP</p> <ul style="list-style-type: none"> • Biases de sélection • Conception • Variables confusionnelles • Traitement à l'aveugle • Méthodes de collecte des données • Retraits et abandons 	<p>Ce modèle de notation a été vérifié et s'avère un outil fiable et valide.</p> <p>Ce système de notation indique une préférence pour les EEAC et les essais cliniques sur échantillon contrôlé, où les notes auront tendance à être plus élevées que pour des études sur échantillon non aléatoire.</p> <p>La meilleure note qui peut être obtenue avec des études par observation est moyenne, et cela se produit seulement quand tous les éléments de l'étude sont très bien conçus.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Ramsay et coll. 2003 États-Unis</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p> <p>MÉTHODES <i>(Modèles à séries chronologiques interrompues (SCI))</i></p>	<p>Examiner de manière critique la qualité méthodologique des modèles de SCI (séries chronologiques interrompues) à l'aide d'études incluses dans deux examens systématiques (un sur les interventions médicales de masse et un sur les stratégies de diffusion et d'application des lignes directrices).</p>	<p>Critères de qualité des modèles de SCI :</p> <ul style="list-style-type: none"> • l'intervention se déroulait indépendamment d'autres changements qui se produisaient au fil du temps; • l'intervention était peu susceptible de toucher la collecte des données; • le résultat était évalué à l'aveugle ou mesuré objectivement; • le résultat était fiable ou mesuré objectivement; • la composition de l'ensemble de données à chaque point abordé englobait au moins 80 % de tous les participants à l'étude; • la forme de l'effet d'intervention était précisée au préalable; • le raisonnement sous-tendant le nombre et l'espacement des points de données était décrit; • l'étude était analysée convenablement selon des techniques par séries chronologiques. <p>Constatations</p> <ul style="list-style-type: none"> • En tout, 66 % (n = 38) des études à SCI n'ont pas permis d'empêcher qu'un autre événement se soit produit au moment de l'intervention. • En général, dans les deux examens, on a communiqué les facteurs touchant la collecte des données, le résultat principal et l'intégralité de l'ensemble de données. • Toutes les études étaient jugées efficaces dans le rapport original, mais environ la moitié des études analysées de nouveau n'ont révélé aucun écart statistique significatif. 	<p>Souvent, les modèles à SCI sont souvent analysés inadéquatement, manquent de puissance et sont mal communiqués dans les études sur la mise en œuvre.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
Conn et Rantz 2003 États-Unis OUTILS D'ÉVALUATION CRITIQUE MÉTHODES (Qualité et résultats d'étude)	1. Discuter des manières dont les chercheurs conçoivent la qualité et l'évaluent. 2. Déterminer les liens qui existent entre la qualité et les résultats des études. 3. Établir des stratégies pour gérer la qualité variable des études primaires dans une méta-analyse.	Évaluer la qualité méthodologique <ul style="list-style-type: none"> • Il n'existe aucune norme relative aux variations de qualité dans les études primaires. • Tableau 1 – résumé des composants courants de la qualité des études sur les interventions. • Instruments de mesure de la qualité – il existe plus de 100 échelles de qualité pour les études primaires, mais leur taille, leur composition, leur complexité et leur portée varient radicalement. • Les échelles de mesure de la qualité présentent des problèmes de mise au point et d'application. • La plupart des échelles donnent une note unique et globale de qualité. Seulement quelques-unes renforcent des sous-échelles qui indiquent les forces et les faiblesses. • Il est prouvé qu'il est difficile d'appliquer les instruments uniformément, même dans les EEAC. <p>Rapport entre la qualité et les résultats d'étude :</p> <ul style="list-style-type: none"> • les constatations sont souvent contradictoires; • les constatations sont inégales quant au fait que les études de faible qualité sous-estiment ou surestiment l'ampleur de l'effet comparativement aux études de haute qualité; • les différentes échelles donnent différentes évaluations de la qualité des études, ce qui entraîne un manque d'uniformité dans les efforts visant à établir un lien entre la qualité des études et leurs résultats. <p>Stratégies pour gérer la qualité</p> <ul style="list-style-type: none"> • Établir des niveaux minimums d'inclusion ou exiger que certains attributs de qualité soient présents — ce qui peut signifier d'établir des critères d'inclusion, de choisir de certaines notes limites de qualité a priori; de fixer des limites : exclure les études qui seraient moins rigoureuses est contraire à l'habitude scientifique selon laquelle les données sont examinées et on les laisse parler. • Pondérer l'ampleur de l'effet par les notes de qualité — permet d'inclure diverses études, mais repose sur des mesures de qualité douteuses. • Considérer la qualité comme une question empirique — permet d'examiner les rapports entre la qualité et la taille de l'effet et ainsi de préserver l'objet de la méta-analyse, soit examiner les données systématiquement; limites : les problèmes que présentent les échelles de mesure de la qualité globale risquent de limiter la confiance accordée aux constatations relatives aux mesures de qualité globale. 	Conclusions <ul style="list-style-type: none"> • Les chercheurs combinent de plus en plus les stratégies pour surmonter les limites que présente l'utilisation d'une seule approche. • S'efforcer davantage de créer des mesures valides de la qualité des études primaires améliorera la capacité des méta-analyses d'influencer la recherche et la pratique infirmière.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Linde et coll. 2002 Allemagne ÉTUDES (sur échantillon aléatoire ou non aléatoire)</p>	<p>Déterminer :</p> <ol style="list-style-type: none"> 1. les différences possibles sur le plan des patients, des interventions, des aspects de la qualité indépendants de la conception et des taux de réponse entre les études sur échantillon aléatoire et non aléatoire concernant l'acupuncture pour le traitement des maux de tête chroniques; 2. si les études sur échantillon non aléatoire procurent d'autres renseignements pertinents (soit, les méthodes et les résultats à long terme, les facteurs prévisionnels, les effets négatifs, les taux de réponse). 3. les explications possibles en cas d'écart entre les taux de réponse pour les patients sélectionnés au hasard ou non. 	<p>Critères de qualité</p> <ul style="list-style-type: none"> • La méthode d'échantillonnage est-elle claire et clairement décrite? • Le diagnostic de maux de tête est-il clairement établi? • Les patients sont-ils caractérisés (au moins selon l'âge, le sexe, la durée et la gravité des symptômes)? • La période de base dure-t-elle au moins quatre semaines? • Obtient-on au moins deux résultats cliniques en ce qui touche les maux de tête? • Utilise-t-on un relevé quotidien des maux de tête? • Les co-interventions sont-elles décrites? • Est-ce qu'au moins 90 % des patients inclus ont fait l'objet d'une analyse après traitement? • Est-ce qu'au moins 80 % des patients traités ont fait l'objet d'une analyse lors du suivi précoce (moins de six mois après)? • Est-ce qu'au moins 80 % des patients traités ont fait l'objet d'une analyse lors du suivi tardif (au moins six mois après)? <p>Méthodes et résultats</p> <ul style="list-style-type: none"> • Cinquante-neuf études ont été incluses : 24 essais sur échantillon aléatoire et 35 études sur échantillon non aléatoire (cinq études de cohortes sur échantillon non aléatoire et contrôlé, 10 études prospectives sur échantillon non contrôlé, 10 séries de cas et 10 enquêtes transversales). • Dix des 24 EEAC et 26 des 35 études sur échantillon non aléatoire répondaient à moins de cinq critères de qualité. • En tout, 535 patients ont reçu un traitement d'acupuncture dans les 24 EEAC, comparativement à 2695 dans les 35 études sur échantillon non aléatoire. • En moyenne, les essais sur échantillon aléatoire se sont faits sur un échantillon plus faible, ont répondu à plus de critères de qualité et ont enregistré des taux de réponse moins élevés (0,59; 0,48 - 0,69) c. 0,78 (0,72 - 0,83). • Sur échantillon aléatoire ou non, les études qui répondaient à plus de critères de qualité enregistraient un taux de réponse moins élevé. • La période de suivi pour les études sur échantillon non aléatoire n'était pas nettement plus longue. Sur les trois qui incluaient une analyse des variables prévisionnelles, seulement une faisait état des effets négatifs, et les degrés de généralisabilité n'étaient pas clairs. 	

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>MacLehose et coll. 2000 Royaume-Uni ÉTUDES (EEAC c. études quasi-expérimentales et par observation (QEO))</p>	<p>Comparer les estimations d'efficacité obtenues par des EEAC et des QEO et examiner le rapport entre la qualité méthodologique et l'ampleur des estimations d'efficacité.</p>	<p>1re stratégie : comparer les estimations quant à l'efficacité de toute intervention qui ont été obtenues par des EEAC et des études QEO, où les deux estimations figureraient dans un seul.</p> <p>Quatorze documents qui contenaient 38 comparaisons, dont 13 étaient jugées de qualité supérieure et 25, de faible qualité.</p> <p>Les écarts entre les estimations obtenues par des EEAC et des études QEO concernant l'ampleur de l'effet et la fréquence des résultats pour les groupes ayant fait l'objet d'une intervention et les groupes témoins étaient plus faibles dans le cas des comparaisons de grande qualité que dans ceux de faible qualité.</p> <p>Conclusion</p> <ul style="list-style-type: none"> • Les estimations d'efficacité obtenue par des études QEO peuvent être valides si les facteurs confusionnels importants sont contrôlés. • Les comparaisons de faible qualité ont tendance à donner des estimations d'effet extrêmes dans le cas des études QEO. • Les méthodes et les résultats sont limités par le nombre de documents examinés (faible) et la nature potentiellement non représentative des données probantes examinées. <p>2e stratégie : comparer les estimations quant à l'efficacité d'interventions précises qui ont été obtenues par des EEAC et des études QEO, où les deux estimations figureraient dans des documents différents.</p> <p>Interventions : dépistage par mammographie destiné à réduire la mortalité due au cancer du sein et les suppléments d'acide folique servant à réduire les anomalies du tube neural chez les femmes tentant de tomber enceintes.</p> <p>Trente-quatre documents ont été examinés (17 sur le dépistage par mammographie, 17 sur les suppléments d'acide folique).</p> <p>Huit et quatre documents, respectivement, étaient des EEAC attribués individuellement ou par grappe, cinq et six étaient des essais sur échantillon non aléatoire ou des études de cohortes, et trois et six étaient des études cas-témoin appariées ou non. Deux études, une sur le dépistage par mammographie et une sur les suppléments d'acide folique, reposaient sur un autre modèle.</p> <p>Les notes de qualité totales étaient plus basses pour les études de cohortes et cas-témoin que pour les EEAC; et les notes pour les études de cohortes étaient nettement plus basses que pour les études cas-témoin.</p> <p>La méta-régression des attributs d'étude par rapport aux estimations de risque relatif n'a révélé aucun lien entre l'ampleur de l'effet et la qualité de l'étude, peu importe l'intervention.</p> <p>Les estimations obtenues par les EEAC et les études de cohortes n'étaient pas nettement différentes, mais les études cas-témoin ont donné des estimations bien différentes lorsqu'il était question de mammographie (avantage supérieur) et d'acide folique (avantage inférieur).</p>	<p>Recommandations</p> <ul style="list-style-type: none"> • Il y aurait lieu d'établir des normes de communication pour les études QEO. • Il faudrait examiner la possibilité de créer des stratégies de recherche nouvelles. • Il faut élaborer des méthodes d'établissement des études qui fournissent une comparaison directe des estimations découlant de données aléatoires et non aléatoires.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Britton 1998 Royaume-Uni ÉTUDES <i>(EEAC c. études sur échantillon non aléatoire)</i></p>	<p>Examiner les questions touchant le processus de randomisation qui peuvent influencer sur la validité des conclusions tirées des méthodes et des résultats des EEAC et des études sur échantillon non aléatoire</p>	<p>Comparaisons antérieures entre des EEAC et des études sur échantillon non aléatoire :</p> <ul style="list-style-type: none"> • Dix-huit documents où l'on comparait directement les résultats d'EEAC et d'études prospectives sur échantillon non aléatoire et les méthodes employées dans les deux cas ont été découverts et analysés. • Les estimations de l'effet du traitement n'ont été invariablement plus élevées ou plus faibles ni pour les EEAC, ni pour les études sur échantillon non aléatoire. • Le genre d'intervention n'a pas semblé exercer d'influence. <p>Exclusions</p> <ul style="list-style-type: none"> • Dans les EEAC, la proportion de sujets admissibles allait de 1 % à 100 %. • Les motifs d'exclusion pouvaient être médicaux (p. ex., risque élevé de situations négatives dans certains groupes) ou scientifiques (sélection axée sur de petits groupes homogènes pour augmenter la précision des effets de traitement estimatifs); de plus, les exclusions globales sont courantes dans les EEAC. <p>Participation</p> <ul style="list-style-type: none"> • Dans la plupart des EEAC, les caractéristiques des personnes admissibles qui n'ont pas participé aux essais sont mal documentées. <p>Adapter en fonction des différences de base</p> <ul style="list-style-type: none"> • Dans les études sur échantillon non aléatoire, l'adaptation en fonction des différences a souvent changé l'ampleur de l'effet du traitement, mais pas de manière significative; ce qui importe, c'est que l'orientation du changement n'était pas constante. 	<p>Conclusions :</p> <ul style="list-style-type: none"> • Les résultats des EEAC et des études sur échantillon non aléatoire et les méthodes employées dans les deux cas ne différaient pas inévitablement • Les données probantes qui existent sont limitées à bien des égards, mais on laisse entendre qu'il est possible de réduire les différences au minimum en s'assurant que les sujets inclus dans chaque genre d'étude sont comparables.
<p>Lemmer et coll. 1999 Royaume-Uni MÉTHODES <i>(ES)</i></p>	<p>Faire état des grands problèmes qu'a présentés un examen systématique de la documentation qui, en l'absence d'EEAC, s'est inspiré d'études théoriques de travaux de recherche fondés sur la prise de décisions et la pratique.</p>	<p>En tout, 164 examens de 82 publications ont été effectués.</p> <p>Une version modifiée du Cochrane Collaboration Protocol a été utilisée :</p> <ul style="list-style-type: none"> • Les bases de données et les mots clés employés pour les stratégies de recherche ont été choisis en consultation avec un bibliothécaire spécialisé en santé et dans des matières connexes. • Les parties omises ont été jugées inappropriées aux fins de comptes rendus qui ne concernaient pas des EEAC, et elles ont été remplacées par des rubriques relatives à une plus vaste gamme de méthodologies. <p>À la fin du protocole, il y avait aussi une partie pour les commentaires des examinateurs sur l'opportunité et l'importance de chaque examen.</p> <p>Une note numérique a été attribuée à chaque article; 8 est la note maximum et 1, la note minimum.</p> <p>Ce barème a servi à indiquer la rigueur méthodologique et la pertinence de chaque article.</p>	<p>Les études qualitatives fournissent peu de renseignements sur la méthodologie, ce qui rend l'évaluation de la qualité difficile. • Cependant, tous les examinateurs ont convenu que certains des articles non cotés comme étant mauvais renfermaient des questions ou des parties importantes dont l'omission aurait nui à l'étude.</p> <p>Il a été difficile de s'entendre sur une note, car chaque examinateur interprétait les articles et les modèles d'étude différemment.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Goldsmith et coll. 2007</p> <p>Royaume-Uni</p> <p>MÉTHODES</p>	<p>Décrire les méthodes d'examen élaborées et les difficultés rencontrées pendant le processus qui consistait à mettre à jour l'examen systématique de données probantes pour influencer les lignes directrices concernant le contenu des données sur les patientes relatif au dépistage du cancer du col de l'utérus.</p>	<p>Les études de femmes dans un programme organisé de dépistage systématique du cancer du col de l'utérus ont été incluses; tous les modèles d'étude, à l'exception des articles d'opinion, ont été inclus.</p> <p>Deux cent trente-trois études ont été récupérées aux fins d'évaluation plus approfondie.</p> <p>Les données ont été extraites de 79 études (52 quantitatives et 27 qualitatives).</p> <p>Trente-deux études ont été indiquées comme étant pertinentes.</p> <p>Treize études qualitatives, sept études descriptives non comparatives, trois EEAC, trois quasi-EEAC, trois études transversales, un EEAC par grappe, une étude cas-témoin rétrospective et une étude de séries chronologiques non comparative ont été inclus dans le compte rendu.</p> <p>La qualité de chaque étude a été évaluée par des listes de contrôle établies : SIGN, CASP, NZGG, UKGCSRO et une dressée par l'examinateur.</p> <p>Les principes fondamentaux de l'approche GRADE ont été appliqués à la synthèse des données probantes quantitatives.</p>	<p>Conclusions</p> <ul style="list-style-type: none"> Ce sont des données qui provenaient de diverses sources qui ont le mieux permis de répondre aux questions d'examen. La recherche qualitative était souvent très pertinente et spécifique à bien des composants des documents d'information sur le dépistage.
<p>DeWalt et coll. 2004</p> <p>États-Unis</p> <p>MÉTHODES</p>	<p>Examiner le rapport entre les résultats pour l'alphabétisation et la santé.</p> <p>Examen d'études – permet d'examiner des études par observation qui indiquaient des données originales mesurant l'alphabétisation par un instrument valide quelconque et un ou plusieurs résultats pour la santé.</p>	<p>Ils ont inclus des études où les résultats touchaient la santé et les services de santé ainsi que la mesure des capacités à lire et à écrire par un instrument valide (West et coll., 2002).</p> <p>Chaque étude a été cotée selon :</p> <ul style="list-style-type: none"> la pertinence de la population à l'étude, la comparabilité des sujets, la validité et la fiabilité de la mesure d'alphabétisation, le maintien de groupes comparables, l'opportunité de la mesure des résultats, l'opportunité de l'analyse statistique, la pertinence du contrôle des variables confusionnelles. 	<p>En moyenne, la qualité des études incluses allait d'assez bonne à bonne.</p> <p>Les auteurs ont constaté que la plupart des études se sont mal attachées au contrôle des variables confusionnelles et au recours à des comparaisons multiples.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Thomson et coll. 2006 Royaume-Uni MÉTHODES</p>	<p>Faire la synthèse de données sur les grands déterminants socio-économiques de la santé et les inégalités en matière de santé qui figurent dans les évaluations menées sur le programme national de régénération au Royaume-Uni..</p>	<p>Ils ont inclus des évaluations qui faisaient état de réalisations en s'inspirant de données provenant de deux secteurs cibles d'un programme national de régénération, ou d'initiatives locales, au Royaume-Uni.</p> <p>Dix-neuf évaluations indiquaient les effets sur la santé ou les déterminants socio-économiques de la santé. Les données ont été synthétisées à partir de 10 évaluations.</p> <p>Des procédures standard d'examen systématique ont été appliquées, y compris :</p> <ul style="list-style-type: none"> • une stratégie de recherche exhaustive, • des critères d'inclusion ou d'exclusion a priori, • l'examen d'articles par deux personnes indépendantes, • l'extraction des données, • la synthèse des données. • 	<p>Les lacunes méthodologiques des études primaires ont mis en péril le processus de synthèse des données.</p> <p>Les auteurs ont laissé entendre que les interventions et les programmes attachés à un cadre théorique (p. ex., la théorie du changement) aideront à constituer les données probantes.</p>
<p>Stein et coll. 2005 États-Unis MÉTHODES <i>(Évaluations des technologies de la santé (ÉTS))</i></p>	<p>Enquêter sur le lien entre la fréquence et les caractéristiques méthodologiques dans un échantillon d'ÉTS.</p>	<p>Les chercheurs ont inclus les examens de quatre interventions publiées dans trois rapports. Les données ont été extraites de rapports publiés ou non par un examinateur et vérifiées par un autre.</p> <p>Divers essais d'analyse de régression ont été entrepris sur les données.</p> <p>Les mesures de résultats d'étude ont été communiquées à différents moments, selon la durée de suivi de toute l'étude.</p> <p>L'examen portait sur :</p> <ul style="list-style-type: none"> • la taille de l'échantillon, • l'approche prospective ou rétrospective, • les organismes à centres multiples ou uniques, • le recrutement consécutif, • l'indépendance de la mesure des résultats, • la durée du suivi, • la date de publication. <p>Constataions</p> <ul style="list-style-type: none"> • Peu de données probantes ont été trouvées quant à l'existence d'un lien entre les caractéristiques méthodologiques et le résultat. • Il n'a pas été révélé que la taille de l'échantillon et l'approche prospective étaient reliées au résultat. 	<p>Tous les résultats concernaient des interventions chirurgicales, ce qui pouvait influencer sur la généralisabilité.</p> <p>Le faible nombre de cas et le nombre limité d'études dans chaque ensemble de séries de cas pourraient limiter la précision et la généralisabilité.</p> <p>Il se peut que les séries de cas soient particulièrement propices aux biais de publication.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
Daiziel et coll. 2005 Royaume-Uni MÉTHODES (Séries de cas dans les ÉTS) ÉTUDES (Séries de cas c. EEAC)	1. Examiner le recours à des séries de cas (SC) dans les comptes rendus d'ÉTS du National Institute for Clinical Excellence (NICE). 2. Examiner systématiquement la documentation méthodologique pour trouver des documents sur la validité d'aspects du modèle des séries de cas. 3. Enquêter sur les caractéristiques et les constatations de séries de cas à l'aide d'exemples du programme d'ÉTS du Royaume-Uni	Examen du recours à des séries de cas dans les ÉTS du NICE : <ul style="list-style-type: none"> • sur les 47 ÉTS réalisées, 14 comprenaient des renseignements provenant de SC; • les critères d'inclusion des SC incluaient la taille de l'étude et la durée du suivi; • les SC à inclure dans les ÉTS, la manière de les utiliser ou celle d'évaluer leur qualité ne faisaient pas consensus. Examen systématique de la documentation méthodologique : <ul style="list-style-type: none"> • une recherche a été menée dans le but de trouver des séries qui évaluaient les aspects modèle, analyse ou qualité des séries de cas par rapport à la validité des études; • aucune étude empirique n'a été trouvée. Enquête sur les caractéristiques et les constatations des modèles de séries de cas : <ul style="list-style-type: none"> • aucun lien n'a été constaté entre la taille de l'échantillon et la fréquence des résultats ou entre la collecte de données prospectives et la fréquence des résultats; • une analyse par domaine (dans divers domaines) a permis de constater un lien significatif entre les études à centres multiples et le résultat, entre la mesure indépendante et la fréquence des résultats et entre la publication précoce et la fréquence des résultats; • la durée du suivi s'est révélée nettement reliée à la fréquence des résultats dans trois analyses. Comparaison entre les séries de cas et les EEAC : <ul style="list-style-type: none"> • comparativement aux données probantes provenant d'EEAC, qui ne révélaient aucune différence entre l'angioplastie coronaire et le pontage aortocoronarien, les estimations de mortalité des séries de cas montraient une augmentation de 1 à 2 % de la mortalité dans le cas du pontage; • dans les cas d'angine récurrente, ni les données provenant des séries de cas, ni celles provenant des EEAC n'ont révélé de différences entre les deux interventions. 	Les études incluses concernaient toutes des interventions chirurgicales, ce qui pourrait limiter leur applicabilité à d'autres interventions ou cadres. L'étude portait sur un faible nombre de cas. Par conséquent, les résultats devraient être considérés avec prudence.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Deeks et coll. 2003 Royaume-Uni</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p> <p><i>(Évaluer la capacité d'adaptation à la clientèle pour contrôler le biais)</i></p>	<p>Songer aux méthodes et aux données probantes connexes en vue d'évaluer le biais dans les études d'intervention sur échantillon non aléatoire.</p>	<p>Huit études visaient à comparer les résultats d'études sur échantillon aléatoire et non aléatoire sur de multiples interventions.</p> <p>En tout, on a relevé 194 outils servant à évaluer les études sur échantillon non aléatoire.</p> <p>Constatations</p> <ul style="list-style-type: none"> • Les résultats des études sur échantillon non aléatoire diffèrent parfois de ceux des études sur échantillon aléatoire sur la même intervention. • Il existe bien des outils d'évaluation de la qualité, mais en évaluant les études, on oublie des domaines de qualité clés. • Les études sur échantillon non aléatoire devraient seulement être entreprises quand il est impossible d'effectuer des EEAC. 	
<p>Katrak et coll. 2004 Australie</p> <p>OUTILS D'ÉVALUATION CRITIQUE</p>	<p>Résumer le contenu, l'intention, la construction et les propriétés psychométriques des outils d'évaluation critique publiés qui existent à l'heure actuelle, afin de relever les éléments communs et de définir leur pertinence à l'égard de la recherche sur la santé.</p>	<p>En tout, 121 outils d'évaluation critique publiés ont été inclus dans l'ES, et ils provenaient de 108 documents.</p> <p>Constatations</p> <ul style="list-style-type: none"> • Les outils étaient spécifiques au modèle d'étude dans une proportion de 87 %, et la plupart d'entre eux avaient été créés pour des études expérimentales (38 % de tous les outils découverts). • Les éléments contenus dans les outils variaient énormément. • Les outils disponibles avaient été créés à l'aide de travaux de recherche empirique précis dans une proportion de 12 % (n = 14 instruments). • Les outils résumaient l'évaluation de la qualité en une note globale numérique dans une proportion de 49 %. • Quelques outils avaient documenté des données probantes quant à la validité de leurs éléments ou à la fiabilité d'utilisation. • Des lignes directrices sur l'administration des outils étaient fournies dans 43 % (n = 52) des cas. 	<p>Conclusions</p> <ul style="list-style-type: none"> • Il n'existe pas d'exemple idéal d'outil d'évaluation critique pour un modèle d'étude ou l'autre, pas plus qu'il n'y a d'outil générique largement reconnu qui puisse être appliqué aussi bien à un type d'étude qu'à un autre. • Les consommateurs d'études devraient choisir les outils avec soin. • Les outils choisis devraient être appuyés par des données probantes publiées à fondement empirique sur le plan de la construction, de la validité des éléments et la fiabilité d'interprétation; et ils devraient être accompagnés de lignes directrices d'utilisation.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Greenhalgh et coll. 2005 Royaume-Uni MÉTHODES (Examen <i>méta-narratif</i>)</p>	<p>Introduire ou décrire la mise au point d'une nouvelle méthode et en faire état : examen méta-narratif.</p>	<p>L'examen méta-narratif a été mis au point comme base méthodologique pour la synthèse de données probantes issues de diverses disciplines. Il est particulièrement solide comme méthode de synthèse quand :</p> <ul style="list-style-type: none"> • la portée d'un projet est large et la documentation, variée; • différents groupes de scientifiques ont posé différentes questions et utilisé divers modèles d'étude pour s'attaquer à un problème commun; • les documents de « qualité » sont définis par différentes caractéristiques dans diverses documentations; • il n'existe aucun processus universellement reconnu pour rassembler les différents travaux. <p>L'étude comprenait une étape de structuration systématique visant à recueillir et à comparer les différentes trames générales concernant l'essor et le déclin de la diffusion, ce qui est jugé pertinent à l'égard de la question de recherche globale.</p> <p>Phases de l'examen méta-narratif :</p> <ul style="list-style-type: none"> • planification • dépouillement • structuration • évaluation • synthèse • recommandations <p>Cinq grands principes sous-tendaient cette technique méta-narrative : le pragmatisme, le pluralisme, l'historicité, la contestation et l'examen par les pairs.</p>	<p>La méthode devrait être mise à l'essai de manière prospective.</p> <p>Il y aurait lieu d'étudier plus à fond son apport à l'économie mixte de méthodes pour l'examen systématique des données probantes complexes.</p>
<p>Whittemore et Knaff 2005 États-Unis MÉTHODES (Examen <i>par intégration</i>)</p>	<p>Faire la distinction entre la méthode d'examen par intégration et les autres méthodes d'examen et proposer des stratégies méthodologiques propres à la méthode par intégration pour augmenter le degré de rigueur.</p>	<p>Enjeux décrits et stratégies employées pour augmenter le degré de rigueur de la méthode d'examen par intégration en soins infirmiers :</p> <ul style="list-style-type: none"> • objet d'examen et variables d'intérêt bien spécifiés — aider à opérationnaliser avec précision les variables et à extraire les données appropriées des sources primaires; • stratégies de dépouillement documentaire bien définies — établir le nombre maximum de sources primaires admissibles à l'aide d'au moins deux ou trois stratégies; • étape d'évaluation des données — comme la méthode d'examen par intégration comprend diverses sources primaires, évaluer la qualité devient de plus en plus complexe; le mode d'évaluation variera selon le cadre d'échantillonnage; dans un examen ou le cadre d'échantillonnage est varié et englobe des sources empiriques et théoriques, il se peut qu'il convienne d'adopter une approche semblable à la recherche historique pour évaluer la qualité; • réduction des données — diviser les sources primaires en sous-groupes suivant une certaine logique pour faciliter l'analyse; • présentation et comparaison des données, établissement de conclusions et vérification, et exposé. 	<p>La méthode d'examen par intégration permet de combiner diverses méthodologies (p. ex., les études expérimentales et non expérimentales) et peut jouer un plus grand rôle dans la pratique fondée sur des données probantes.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Norris et Atkins 2005 États-Unis MÉTHODES (ES)</p>	<p>Examiner le recours à des études sur échantillon non aléatoire dans les comptes rendus de centres de pratique fondée sur des données probantes, touchant des questions d'efficacité des interventions de traitement.</p>	<p>Sur les 107 comptes rendus de centres de pratique fondée sur des données probantes qui ont été publiés de février 1999 à septembre 2004, 78 visaient à examiner au moins une question d'efficacité d'une intervention clinique.</p> <p>Quarante-neuf des comptes rendus incluaient des données probantes provenant de modèles d'étude autres que les EEAC; ces comptes rendus portaient sur la pharmacothérapie, les appareils médicaux, la chirurgie et les interventions complémentaires, de rechange ou comportementales.</p> <p>Défis que représente le recours à des études sur échantillon non aléatoire dans les examens systématiques :</p> <p>la terminologie employée pour décrire différents modèles d'étude sur échantillon non aléatoire dans la documentation sur les travaux cliniques et les comptes rendus des centres de pratique fondée sur des données probantes n'est pas uniforme;</p> <p>il n'existe aucune ligne directrice établie quant au moment où l'inclusion d'études sur échantillon non aléatoire dans les examens systématiques peut ou devrait être envisagée, et aux modèles d'étude à considérer;</p> <p>il est difficile d'évaluer la qualité des études sur échantillon non aléatoire;</p> <p>sur les 49 comptes rendus qui incluaient des modèles d'étude sur échantillon non aléatoire :</p> <ul style="list-style-type: none"> • 12 (25 %) n'évaluaient pas la qualité de l'étude; • 16 % reposaient sur une liste de contrôle ou un système de notation; • 10 % étaient adaptés à partir d'un texte déjà publié; • le reste (49 %) reposait sur des textes que les examinateurs avaient rédigés eux-mêmes. 	<p>Recommandations pour les auteurs d'examen systématique :</p> <ul style="list-style-type: none"> • évaluer la disponibilité d'EEAC avant de déminer les critères d'inclusion finals; • peser le pour et le contre des modèles par EENC; • fournir un raisonnement pour les décisions d'inclure ou non des modèles d'étude en particulier; • évaluer les domaines importants de qualité d'étude; • se demander en quoi inclure divers modèles d'étude pourrait influencer sur les conclusions; • se demander en quoi la qualité des études et l'ensemble des données probantes pourraient influencer sur les conclusions. <p>Recommandations pour les chercheurs :</p> <ul style="list-style-type: none"> • réduire au minimum les sources de biais, peu importe le modèle d'étude; • examiner les effets de diverses sources de biais sur les résultats mesurés; • employer à une terminologie uniforme pour le modèle d'étude; • prévoir des stratégies de recherche exhaustives pour les modèles d'étude sur échantillon non aléatoire.

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Jackson et Walters 2005</p> <p>Australie</p> <p>MÉTHODES (ES)</p>	<p>Fournir des recommandations aux examinateurs sur les questions à régler dans un examen systématique en santé publique et donner indirectement des conseils aux chercheurs sur les exigences de communication rattachées aux études primaires concernant la production d'examens systématiques de grande qualité.</p>	<p>Afin de s'assurer que les examens répondent aux besoins des utilisateurs, former un groupe consultatif (dont les membres devraient bien connaître le sujet, l'utilité d'inclure le point de vue de responsables de l'élaboration de politiques, de bailleurs de fonds, de praticiens et d'utilisateurs potentiels).</p> <p>Enjeux auxquels s'attacher quand il s'agit d'examiner des interventions de santé publique :</p> <ul style="list-style-type: none"> • inclusion de modèles d'étude • recherche de documentation en santé publique • évaluation de la qualité • cadres théoriques d'intervention • intégrité des interventions • hétérogénéité • intégration d'études qualitatives et quantitatives • éthique et inégalités – interventions de santé permettant de réduire les inégalités et d'améliorer les personnes marginalisées ou défavorisées 	<ul style="list-style-type: none"> • Recommandation • Utiliser le <i>Quality Assessment Tool for Quantitative Studies mis au point</i> par Thomas et coll., 2004.
<p>Atkins et Di-Giuseppe 1998</p> <p>États-Unis</p> <p>MÉTHODES (Recherche sur les soins de santé préventifs)</p>	<p>En s'inspirant des expériences vécues par le US Preventive Services Task Force, exposer dans leurs grandes lignes certains grands domaines où il faut effectuer des recherches afin de définir la manière appropriée d'utiliser des services de prévention précis (p. ex., les examens de dépistage, les interventions de consultation, les vaccinations, la chimioprophylaxie).</p>	<p>Études par observation :</p> <ul style="list-style-type: none"> • aident à établir des liens dans l'enchaînement de causalité; • aident à la comprendre l'histoire naturelle de la maladie et à définir les facteurs de risque, ainsi qu'à mesurer le respect des traitements et les effets négatifs de ceux-ci; • aident à établir l'exactitude des tests de diagnostic; • aident à évaluer l'efficacité des interventions; • possèdent un biais inhérent. 	<p>Les recommandations fondées seulement sur les données probantes obtenues par observation exigent des études de grande qualité qui révèlent des effets constants et importants, de préférence.</p> <p>Quand on dispose uniquement de quelques études par observation de qualité adéquate ou l'ampleur des effets est modeste ou incohérente, d'autres études sur populations différentes s'avèreraient un ajout précieux à la documentation.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Harden et coll. 2004 Royaume-Uni</p> <p>MÉTHODES (Études de « points de vue »)</p>	<p>Décrire les méthodes mises au point pour examiner les études sur les points de vue et les expériences des gens (études de « points de vue »), en plus d'effectuer des essais, dans une série d'exams sur la santé mentale, l'activité physique et la saine alimentation des jeunes.</p>	<p>Deux types d'études ont été examinés :</p> <ol style="list-style-type: none"> 1. les études sur les interventions – définir les interventions efficaces, inefficaces et nocives; 2. les études sur les non-interventions – décrire les facteurs associés à la santé mentale, à l'activité physique et à la saine alimentation. <p>Études de « points de vue » :</p> <ul style="list-style-type: none"> • trente-cinq répondaient aux critères d'inclusion. <p>Les études variaient selon les méthodes employées — bon nombre d'entre elles ne pouvaient être classées facilement comme étant « qualitatives » ou « quantitatives ».</p> <p>La plupart d'entre elles ne répondaient pas aux sept normes fondamentales de communication de la méthodologie qui sont utilisées dans un outil d'évaluation de la qualité nouvellement mis au point.</p> <p>Voici les avantages que peut apporter le fait de réunir des études de points de vue de manière systématique :</p> <ul style="list-style-type: none"> • obtenir un plus large éventail de points de vue et une meilleure compréhension des questions de santé publique d'après le point de vue de ceux qui sont visés par les interventions; • mieux réfléchir aux méthodes d'étude qui peuvent déformer ou omettre le point de vue des gens; • créer de meilleures occasions pour que les points de vue et les expériences des gens influencent les politiques visant à promouvoir leur santé. 	<p>Nouveau cadre indiqué dans une publication sub-séquente (Oliver et coll., 2005)</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
<p>Wong et Raabe 1996 États-Unis MÉTHODES <i>(Méta-examen)</i></p>	<p>Fournir quelques lignes directrices de base pour que les non-épidémiologistes évaluent les méta-analyses dans les études de cohortes professionnelles.</p>	<p>Certains problèmes et limites que présentent les examens qualitatifs traditionnels :</p> <ul style="list-style-type: none"> la sélection des études n'était normalement pas assujettie à des critères préétablis; les facteurs de pondération attribués à chaque étude étaient habituellement subjectifs; aucun résumé quantitatif des estimations de risques n'était réalisé. <p>Constataions</p> <ul style="list-style-type: none"> Au lieu de remplacer les examens qualitatifs, la méta-analyse quantitative devrait faire partie de l'évaluation globale. Le terme « méta-examen » est proposé pour que l'accent soit mis sur l'importance des composants qualitatifs et quantitatifs d'un processus d'examen exhaustif, ce qui signifie qu'il englobe une méta-analyse. <p>Étapes fondamentales à suivre pour procéder un méta-examen :</p> <ol style="list-style-type: none"> définir la question de recherche; procéder à un examen documentaire exhaustif; créer des critères d'inclusion a priori; effectuer un examen qualitatif traditionnel; effectuer une méta-analyse quantitative; intégrer un examen qualitatif traditionnel avec une méta-analyse quantitative; appliquer les critères de causalité à l'interprétation. 	<p>Avantages du méta-examen :</p> <ul style="list-style-type: none"> utile lorsqu'il s'agit de choisir les études, d'organiser, de présenter et de résumer des méthodes ou des résultats provenant d'études individuelles; peut aussi servir à déceler de l'hétérogénéité parmi les études. <p>Principaux avantages à effectuer une méta-analyse :</p> <ul style="list-style-type: none"> augmenter la puissance statistique (en particulier quand les études initiales sont de faible envergure); fournir une estimation bien pondérée du risque associé aux résumés; prendre en compte l'uniformité de l'étude et des méthodes ou résultats; réduire au minimum le problème des comparaisons multiples; permettre d'examiner les données pour déterminer le degré d'hétérogénéité.
<p>Sandelowski et coll. 2007 États-Unis MÉTHODES <i>(Méta-résumé)</i></p>	<p>S'attaquer au problème associé à la gestion des différences qui existeraient entre les études qualitatives et quantitatives, et promouvoir le recours au méta-résumé qualitatif comme technique utile pour faire la synthèse des constataions descriptives de nature qualitative et quantitative.</p>	<p>Quarante-deux comptes rendus (35 articles de revue, six thèses ou dissertations non publiées et un rapport technique) ont été inclus.</p> <p>Sur les quarante-deux comptes rendus, 12 étaient des études qualitatives, trois, des études sur les interventions, un, une étude sur les méthodes mélangées et 26, diverses études quantitatives par observation.</p> <p>Caractéristiques des méta-résumés qualitatifs :</p> <p>stratégie de regroupement à orientation quantitative à l'égard de la synthèse des travaux de recherche;</p> <ul style="list-style-type: none"> l'extraction, le groupage et le formatage des constataions, ainsi que le calcul de l'ampleur de la fréquence et de l'effet; permettent de faire la synthèse des méthodes et résultats d'enquêtes qualitatives et quantitatives sur les réponses obtenues à partir de procédures de collecte et d'analyse de données semblables; permettent d'effectuer des analyses a posteriori du lien qui existe entre les comptes rendus et les constataions. 	<p>Contrairement à la plupart des méthodes d'examen systématique, cette étude a tendance à permettre l'inclusion plutôt que de l'exclusion des études.</p>

Étude	Objet	Méthodes ou résultats	Commentaires ou enjeux
Jefferson et Demichelli 1999 Royaume-Uni ÉTUDES <i>(expérimentales c. non expérimentales)</i>	Examiner le rapport qui existe entre les modèles d'étude expérimentaux et non expérimentaux en vaccinologie.	Les auteurs ont décrit des études expérimentales et non expérimentales et les approches relatives à chaque type. Ils ont évalué la capacité de chaque modèle d'étude à permettre l'analyse de quatre aspects du rendement des vaccins, soit l'immunogénicité, la durée de l'immunité accordée, l'incidence et la gravité des effets secondaires et le nombre d'infections prévenues par vaccination.	Il y a lieu d'appliquer les modèles non expérimentaux quand : <ul style="list-style-type: none"> • une expérience est impossible parce que l'évaluation concerne l'efficacité du vaccin; • une expérience est inutile (p. ex., dans le cas de l'évaluation du vaccin contre la varicelle); • une expérience est inappropriée (la population visée par l'essai n'est pas assez vaste pour permettre de déceler un cas ou un résultat, une hésitation à participer ou un refus de le faire, les obstacles éthiques, juridiques ou politiques); • l'efficacité individuelle est mesurée du point de vue de la faible fréquence des situations négatives; • les interventions préviennent de rares cas; • l'efficacité du vaccin pour la population doit être mesurée du point de vue des incidences à long terme, rares et graves de maladie.
Roifstein et Lau-pacis 2004 Canada ÉTUDES <i>(ÉTS c. ES)</i>	Établir clairement les grandes différences entre les évaluations des technologies de la santé (ÉTS) et les examens systématiques (ES).	Les chercheurs ont interrogé 17 auteurs ou utilisateurs d'ÉTS. Ils ont défini sept domaines de différences entre les ÉTS et les ES : <ol style="list-style-type: none"> 1. les normes méthodologiques — les ÉTS peuvent inclure de la documentation de mauvaise qualité méthodologique si un sujet est important pour les décideurs; 2. la reprise d'études antérieures — les examens systématiques n'ont pas à être repris si les études antérieures étaient de grande qualité ou quand il n'existait aucune nouvelle donnée probante de grande qualité; dans le cas des ÉTS, il faut souvent reprendre des études pour défendre les conclusions du rapport; 3. le choix des sujets — les sujets sont plus axés sur les politiques avec les ÉTS, alors que les examens systématiques tendent à être orientés par des questions d'efficacité; 4. l'inclusion de spécialistes du contenu (dans les examens systématiques) et de responsables de l'élaboration des politiques (dans les ÉTS) comme auteurs; 5. l'inclusion d'évaluations économiques, dans les ÉTS; 6. la formulation de recommandations d'orientation, dans les ÉTS; 7. la diffusion du rapport, réalisée plus souvent activement dans le cas des ÉTS 	Les ÉTS ne concernent pas seulement l'évaluation des données probantes scientifiques.

Références bibliographiques

- Scottish Intercollegiate Guidelines Network 2001–2005. (2001). *SIGN 50: A guideline developers' handbook*, récupéré le 1er février 2009 à partir du site <http://www.sign.ac.uk/guidelines/full-text/50/index.html>
- Critical Appraisal Skills Programme (CASP) and evidence-based practice. (2005). *Critical Appraisal Skills Programme: Making sense of evidence - CASP appraisal tools*, Oxford, Public Health Resource Unit, Milton Keynes Primary Care NHS Trust, récupéré le 1er février 2009 à partir du site <http://www.phru.nhs.uk/Pages/PHD/resources.htm>
- Atkins, D., Briss, P. A., Eccles, M., Flottorp, S., Guyatt, G. H., Harbour, R. T. et coll. (2005). Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system, *BMC Health Services Research*, 5, 25
- Atkins, D. et DiGuseppi, C. G. (1998). Broadening the evidence base for evidence-based guidelines. A research agenda based on the work of the U.S. Preventive Services Task Force, *American Journal of Preventive Medicine*, 14, 335-344
- Benson, K. et Hartz, A. J. (2000). A comparison of observational studies and randomized, controlled trials, *The New England Journal of Medicine*, 342, 1878-1886
- Britton, A., McKee, M., Black, N., McPherson, K., Sanderson, C. et Bain, C. (1998). Choosing between randomized and non-randomized studies: A systematic review, *Health Technology Assessment*, 2, 1-124
- Chou, R. et Helfand, M. (2005). Challenges in systematic reviews that assess treatment harms, *Annals of Internal Medicine*, 142, 1090-1099
- Conn, V. S. et Rantz, M. J. (2003). Focus on research methods. Research methods: Managing primary study quality in meta-analysis, *Research in Nursing & Health*, 26, 322-333
- Dalziel, K., Round, A., Stein, K., Garside, R., Castelnovo, E. et Payne, L. (2005). Do the findings of case series studies vary significantly according to methodological characteristics? *Health Technology Assessment*, 9, iii-iv, 1
- Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F. et coll. (2003). Evaluating non-randomized intervention studies. *Health Technology Assessment*, 7, 1-173.
- DeWalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N. et Pignone, M. P. (2004). Literacy and health outcomes: A systematic review of the literature, *Journal of General Internal Medicine*, 19, 1228-1239
- DiCenso, A., Prevost, S., Benefield, L., Bingle, J., Ciliska, D., Driever, M. et coll. (2004). Evidence-based nursing: Rationale and resources, *Worldviews on Evidence-Based Nursing*, 1, 69-75
- Downs, S. H. et Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions, *Journal of Epidemiology and Community Health*, 52, 377-384
- Goldsmith, M. R., Bankhead, C. R. et Austoker, J. (2007). Synthesising quantitative and qualitative

research in evidence-based patient information, *Journal of Epidemiology and Community Health*, 61, 262-270

- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. et Peacock, R. (2005). Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review, *Social Science & Medicine*, 61, 417-430
- Greer, N., Mosser, G., Logan, G. et Halaas, G. W. (2000). A practical approach to evidence grading, *Joint Commission Journal on Quality Improvement*, 26, 700-712
- Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G. et coll. (2004). Applying systematic review methods to studies of people's views: An example from public health research, *Journal of Epidemiology and Community Health*, 58, 794-800
- Jackson, N. et Waters, E. (2005). Criteria for the systematic review of health promotion and public health interventions, *Health Promotion International*, 20, 367-374
- Jefferson, T. et Demicheli, V. (1999). Relation between experimental and non-experimental study designs. HB vaccines: A case study, *Journal of Epidemiology and Community Health*, 53, 51-54
- Katrak, P., Bialocerkowski, A. E., Massy-Westropp, N., Kumar, S. et Grimmer, K. A. (2004). A systematic review of the content of critical appraisal tools, *BMC Medical Research Methodology*, 4, 22
- Khan, K., ter Riet, G., Popay, J., Nixon, J. et Kleijnen, J. (2001). Stage II: Conducting the review. Phase 5: Study quality assessment, dans : K. S. Khan, G. ter Riet, J. Glanville et coll. *Undertaking systematic reviews of research on effectiveness. CRD's guidance for carrying out or commissioning reviews 2nd Edition* (pp. 1-20), York : NHS Centre for Reviews and Dissemination, University of York
- Lemmer, B., Grellier, R. et Steven, J. (1999). Systematic review of non-random and qualitative research literature: Exploring and uncovering an evidence base for health visiting and decision making, *Qualitative Health Research*, 9, 315-328
- Lethaby, A., Wells, S. et Furness, S. (2001). *Handbook for the preparation of explicit evidence-based clinical practice guidelines*. Auckland, Nouvelle-Zélande : New Zealand Guidelines Group, Effective Practice Institute of the University of Auckland
- Linde, K., Scholz, M., Melchart, D. et Willich, S. N. (2002). Should systematic reviews include non-randomized and uncontrolled studies? The case of acupuncture for chronic headache, *Journal of Clinical Epidemiology*, 55, 77-85
- MacLehose, R. R., Reeves, B. C., Harvey, I. M., Sheldon, T. A., Russell, I. T. et Black, A. M. (2000). A systematic review of comparisons of effect sizes derived from randomized and non-randomized studies. *Health Technology Assessment*, 4, 1-154
- Margetts, B. M., Thompson, R. L., Key, T., Duffy, S., Nelson, M., Bingham, S. et coll. (1995). Development of a scoring system to judge the scientific quality of information from case-control and cohort studies of nutrition and disease, *Nutrition and Cancer*, 24, 231-239
- McIntosh, H. M., Woolacott, N. F. et Bagnall, A. M. (2004). Assessing harmful effects in systematic

reviews, *BMC Medical Research Methodology*, 4, 19

- Norris, S. et Atkins, D. (2005). Challenges in using nonrandomized studies in systematic review of treatment interventions, *Annals of Internal Medicine*, 142, 1112-1119
- Ogilvie, D., Egan, M., Hamilton, V. et Petticrew, M. (2005). Systematic reviews of health effects of social interventions: 2. Best available evidence: How low should you go? *Journal of Epidemiology and Community Health*, 59, 886-892
- Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. et coll. (2005). An emerging framework for including different types of evidence in systematic reviews for public policy, *Evaluation*, 11, 428-446
- Ramsay, C. R., Matowe, L., Grilli, R., Grimshaw, J. M. et Thomas, R. E. (2003). Interrupted time series designs in health technology assessment: Lessons from two systematic review of behavior change strategies, *International Journal of Technology Assessment in Health Care*, 19, 613-623
- Rangel, S. J., Kelsey, J., Colby, C. E., Anderson, J. et Moss, R. L. (2003). Development of a quality assessment scale for retrospective clinical studies in pediatric surgery, *Journal of Pediatric Surgery*, 38, 390-396
- Rotstein, D. et Laupacis, A. (2004). Differences between systematic reviews and health technology assessments: A trade-off between the ideals of scientific rigor and the realities of policy making, *International Journal of Technology Assessment in Health Care*, 20, 177-183
- Sandelowski, M., Barroso, J. et Voils, C. I. (2007). Using qualitative meta-summary to synthesize qualitative and quantitative descriptive findings, *Research in Nursing & Health*, 30, 99-111
- Slim, K., Nini, E., Forestier, D., Kwiatkowski, F., Panis, Y. et Chipponi, J. (2003). Methodological index for non-randomized studies (minors): Development and validation of a new instrument, *ANZ Journal of Surgery*, 73, 712-716
- Spencer, L., Ritchie, J. et Lewis, J. (2004). *Quality in qualitative evidence: A framework for assessing research evidence*, (2e éd.) Londres : Government Chief Social Researcher's Office
- Stein, K., Dalziel, K., Garside, R., Castelnuovo, E. et Round, A. (2005). Association between methodological characteristics and outcome in health technology assessments which included case series, *International Journal of Technology Assessment in Health Care*, 21, 277-287
- Steuten, L. M., Vrijhoef, H. J., van-Merode, G. G., Severens, J. L. et Spreeuwenberg, C. (2004). The health technology assessment – Disease management instrument reliably measured methodologic quality of health technology assessments of disease management, *Journal of Clinical Epidemiology*, 57, 881-888
- Thomas, B. H., Ciliska, D., Dobbins, M. et Micucci, S. (2004a). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions, *Worldviews on Evidence-Based Nursing*, 1, 176-184
- Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R. et coll. (2004b). Integrating qualitative research with trials in systematic reviews, *British Medical Journal*, 328, 1010-1012

- Thomson, H., Atkinson, R., Petticrew, M. et Kearns, A. (2006). Do urban regeneration programmes improve public health and reduce health inequalities? A synthesis of the evidence from UK policy and practice (1980–2004), *Journal of Epidemiology and Community Health*, 60, 108-115
- West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F. et coll. (2002). Systems to rate the strength of scientific evidence, *Evidence Report – Technology Assessment (Summary)*, 1-11
- Whittemore, R. et Knaf, K. (2005). The integrative review: Updated methodology, *Journal of Advanced Nursing*, 52, 546-553
- Wong, O. et Raabe, G. K. (1996). Application of meta-analysis in reviewing occupational cohort studies, *Occupational and Environmental Medicine*, 53, 793-800
- Zaza, S., Wright-De Aguero, L. K., Briss, P. A., Truman, B. I., Hopkins, D. P., Hennessy, M. H. et coll. (2000). Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services, *American Journal of Preventive Medicine*, 18, 44-74