

Améliorer les pratiques et l'organisation des soins : méthodologie des revues systématiques

Vincent Zaugg, Virginie Savoldelli, Brigitte Sabatier, Pierre Durieux

DANS **SANTÉ PUBLIQUE** 2014/5 Vol. 26 , PAGES 655 À 667

ÉDITIONS **S.F.S.P.**

ISSN 0995-3914

DOI 10.3917/spub.145.0655

Date de mise en ligne : 05/12/2014

Article disponible en ligne à l'adresse

<https://stm.cairn.info/revue-sante-publique-2014-5-page-655?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour S.F.S.P..

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

Améliorer les pratiques et l'organisation des soins : méthodologie des revues systématiques

Improving practice and organisation of care: methodology for systematic reviews

Vincent Zaugg¹, Virginie Savoldelli^{1,2}, Brigitte Sabatier^{1,3}, Pierre Durieux^{1,3}

↳ Résumé

Le nombre d'études d'interventions visant à améliorer la qualité des soins est en constante augmentation, si bien qu'il est difficile d'accéder à toute l'information disponible sur un sujet. La revue systématique est un outil permettant de fournir rapidement une information exhaustive et objective aux professionnels de la santé. Cet article a pour objectif de détailler les étapes de la réalisation d'une revue systématique : définition de la question de recherche, recherche et sélection des études, extraction et analyse des données, évaluation de la qualité méthodologique des études, synthèse des résultats. Les interventions destinées à améliorer les pratiques et l'organisation des soins présentent certaines spécificités qui conditionnent la méthodologie de la revue systématique. Il faut par exemple prendre en compte la variabilité importante qui existe souvent entre les populations, les organisations et les interventions d'une étude à l'autre, ce qui rend difficile la synthèse des résultats sous forme de méta-analyse. La connaissance des particularités des revues systématiques d'interventions est indispensable, que ce soit pour réaliser une synthèse de la littérature de qualité, ou pour évaluer le niveau de preuve d'une revue systématique publiée.

Mots-clés : Études d'intervention ; Qualité des soins de santé ; Pratique professionnelle ; Évaluation des résultats et des processus en soins de santé ; Organisation des soins ; Revue de la littérature.

↳ Summary

The number of intervention studies designed to improve quality of care is increasing exponentially, making it difficult to access all available information on a given subject. Systematic reviews are tools that provide health professionals with comprehensive and objective information. This article describes the main phases of a systematic review: formulating the research question, search and selection of studies, data extraction and analysis, assessment of the methodological quality of studies, and synthesis of the results. Interventions designed to improve professional practices and organisation of care have specific characteristics that determine the methodology of systematic reviews. For example, the often substantial heterogeneity between populations, organisations, and intervention settings among studies must be taken into account, which makes meta-analysis more difficult. Knowledge on specific features of systematic reviews designed to improve quality of care is essential to ensure a good review of the literature, or to evaluate the level of evidence of published systematic reviews.

Keywords: Intervention Studies; Quality of Health Care; Professional Practice; Outcome and Process Assessment (Health Care); Organisation of Care; Review.

¹ Hôpital Européen Georges Pompidou – 20 rue Leblanc – 75015 Paris.

² Université Paris Sud – Faculté de pharmacie – Département de pharmacie clinique – EA123 Barrières physiologiques et réponses thérapeutiques.

³ INSERM – Institut des cordeliers – UMRS 1138 eq22 – Centre Cochrane français.

Introduction

Une **revue systématique** consiste en une synthèse de la littérature scientifique en réponse à une question précise. Elle utilise des méthodes explicites de recherche, de sélection et d'analyse des données. Des méthodes statistiques peuvent ou non être utilisées pour synthétiser les résultats des études. Lorsque de telles méthodes sont employées, on parle de **méta-analyse** [1]. Une revue systématique est donc une vraie méthode de recherche, et ne doit pas être confondue avec une **revue générale** dans laquelle la recherche bibliographique n'est en général pas exhaustive et qui représente plus l'opinion d'un expert ou d'un groupe d'experts.

Les revues systématiques permettent aux professionnels de la santé d'accéder à une information valide et contribuent ainsi à orienter les décisions. Les lecteurs doivent en connaître les règles méthodologiques et être en mesure d'en évaluer la qualité. Les revues systématiques se sont considérablement développées ces dernières années pour évaluer les stratégies thérapeutiques et se sont étendues plus récemment à d'autres domaines, diagnostique et pronostique par exemple. Dans le champ de la qualité et de l'organisation des soins, où le nombre d'études d'intervention publiées est en constante augmentation, les revues systématiques ont également leur place.

Cet article a pour objectif de décrire les étapes et les spécificités méthodologiques d'une revue systématique d'études d'interventions, en se fondant notamment sur les recommandations du **groupe EPOC** [2] (*Effective Practice and Organisation of Care*) de la collaboration Cochrane. Ce groupe publie des revues systématiques sur les interventions destinées « à promouvoir la pratique professionnelle la plus efficace », qu'il s'agisse d'interventions sur les professionnels de santé ou sur le financement ou l'organisation des soins.

Rédaction et enregistrement du protocole

La réalisation d'une revue systématique constitue un travail complexe qui répond aux mêmes règles que toute étude de recherche. Sa première étape est donc la conception et la rédaction d'un protocole. Celui-ci permet de limiter les biais et d'assurer la meilleure objectivité à la revue systématique. Pour sa rédaction, les auteurs peuvent s'appuyer sur les recommandations de la **collaboration Cochrane** [1].

L'enregistrement préalable du protocole de la revue systématique est fortement recommandé afin d'assurer la transparence de l'étude. Cela permet aux lecteurs de détecter d'éventuels biais de report (différences entre le protocole initial et la revue publiée) ou de mettre en évidence des protocoles enregistrés mais non publiés. Les revues systématiques peuvent être enregistrées sur les registres d'essais cliniques comme *clinicaltrials.gov*, ou encore sur la base **PROSPERO** [3], spécialisée dans l'enregistrement des revues systématiques. Lors de l'enregistrement du protocole, un numéro d'enregistrement est fourni aux auteurs. Une fois cité dans l'article publié, il permet d'établir le lien entre la revue systématique et son protocole.

Définition de la question de recherche et des critères d'éligibilité

Définition de la question de recherche

La **question de recherche** doit être formulée de façon claire et précise, car elle conditionne la suite de la revue. On peut pour cela s'inspirer d'autres revues, des recommandations du groupe EPOC de la collaboration Cochrane et de la **méthode PICOT** (*Participants, Interventions, Comparisons, Outcomes, Time*). La question peut être ciblée (par exemple, évaluation de rappels informatiques visant à améliorer la prescription d'antibiotiques) ou plus large (par exemple, évaluation des interventions destinées à améliorer l'usage des antibiotiques). Du choix de la question de recherche dépend la pertinence de la revue. Ainsi, une question trop ciblée peut ne pas être généralisable à d'autres systèmes de soins, alors qu'une question trop large risque d'entraîner des difficultés de synthèse et d'interprétation des résultats. Lors de la définition de la question de recherche, il convient également de vérifier que celle-ci n'a pas déjà été résolue récemment, grâce à une recherche approfondie de la littérature. Des bases de données relatives aux travaux publiés dans le champ de l'organisation des soins ont été développées, comme KT plus (<http://plus.mcmaster.ca/kt/>) et pdq Evidence (www.pdq-evidence.org/fr), et peuvent être consultées à cet effet.

Définition des critères d'éligibilité

Une fois la question de recherche définie, il est nécessaire de préciser les **critères d'éligibilité des études**.

Ces spécifications sur lesquelles les auteurs se basent pour sélectionner les études à inclure dans la revue sont une combinaison du type d'études, des participants, du type d'interventions et des critères de jugement. Cette étape est là encore fondamentale et doit être discutée en groupe avec les personnes qui seront impliquées dans le travail. Par exemple, si l'on souhaite réaliser une revue sur les interventions destinées à améliorer le suivi des malades diabétiques, il faut définir ce que l'on entend par malade diabétique, les tranches d'âge prises en compte, le contexte de soins (prise en charge ambulatoire ou à l'hôpital par exemple), les critères de jugement que l'on souhaite analyser dans les études, et la durée de suivi minimale des patients. Si l'on a défini la mortalité à cinq ans comme critère de jugement principal, les études ne comportant pas ce critère seront exclues.

Il faut également définir si l'on inclut uniquement les essais contrôlés randomisés, ou si d'autres types d'études sont éligibles. Ainsi, le groupe EPOC de la collaboration Cochrane considère d'autres modèles permettant de produire des résultats avec un niveau de preuve acceptable, les études dites quasi expérimentales. Il s'agit des séries chronologiques et des études avant/après avec groupes témoins. Bien que leur niveau de preuve soit moins bon que celui d'essais randomisés, ces études sont parfois la seule

option permettant d'évaluer l'effet d'une intervention. Leur inclusion ou leur exclusion dépendent du type d'intervention évaluée et doivent être discutées entre les auteurs de la revue. L'ensemble des critères d'éligibilité permet d'orienter la stratégie de recherche et d'affiner la sélection des études.

Recherche et sélection des études

Sources de données bibliographiques à consulter

Afin de s'assurer de l'exhaustivité de la recherche, il est indispensable de rechercher les études dans plusieurs bases de données, car aucune base ne les contient toutes. Les trois principales bases de données bibliographiques pour la recherche d'études d'interventions sont **MEDLINE**, **EMBASE**, et **CENTRAL** (*Cochrane Central Register of Controlled Trials*) auxquelles peuvent s'ajouter des bases régionales ou spécifiques à une discipline (tableau I). L'interrogation de chaque base nécessitant une méthodologie particulière, il est important de connaître les guides

Tableau I : Exemples de bases de données pouvant être consultées lors de la réalisation d'une revue systématique d'interventions visant à améliorer la qualité des soins

Base de données	Exemple d'interface de recherche	Vocabulaire contrôlé	Couverture et particularités
MEDLINE	PubMed	MeSH	Base de données produite par la <i>National Library of Medicine</i> (NLM), indexant des revues de biologie et de médecine, principalement anglo-saxonnes.
EMBASE	embase.com	EMTREE	Base de données produite par Elsevier Science, indexant davantage de revues de pharmacologie et de revues européennes que MEDLINE.
CENTRAL	TheCochrane-Library	MeSH, seulement pour les articles provenant de MEDLINE	Base de données produite par la collaboration Cochrane indexant des rapports d'essais contrôlés randomisés à partir de MEDLINE, d'EMBASE, et de recherches manuelles dans des revues et des registres spécialisés.
PASCAL	Ovid	Vocabulaire INIST	Base de données produite par l'Institut de l'Information Scientifique et Technique (INIST), indexant des revues internationales de différentes disciplines, dont près de la moitié sont européennes.
CINAHL	EBSCOhost	Descripteurs CINAHL	Base de données indexant des revues consacrées aux sciences infirmières et paramédicales.
PsycINFO	psycnet.apa.org	Thesaurus of Psychological Index Terms	Base de données produite par l'American Psychological Association (APA), indexant des revues consacrées à la psychologie et à des disciplines connexes (médecine, neuropsychologie, psychiatrie, éducation, sociologie...)
BiblioMab	EPPI-Centre	Mots-clés EPPI	Base de données produite par le <i>Evidence for Policy and Practice Information and Co-ordinating Centre</i> (EPPI-Centre), indexant des études relatives à la santé publique et à la promotion de la santé.

d'utilisation des bases ou de se faire aider par un ou une documentaliste spécialisé(e), afin de s'assurer de la pertinence des recherches. L'utilisation de moteurs de recherche plus généraux, tels que **Google Scholar**, peut aussi permettre d'améliorer l'exhaustivité de la recherche.

Afin d'éviter un **biais de publication**, il est recommandé de réaliser une recherche de la littérature grise [4], c'est-à-dire non contrôlée par un éditeur commercial. Il peut s'agir d'actes de conférence, de publications gouvernementales ou industrielles, d'avis d'experts ou de divers rapports. Cette recherche peut être effectuée à partir de bases de données spécialisées, telles qu'**OpenGrey** [5], développée par l'Institut de l'information scientifique et technique (INIST), ou **The Grey Literature Report** [6], géré par l'Académie de médecine de New York (NYAM).

Les études doivent également être recherchées dans les registres d'essais cliniques. Le registre le plus connu est le site *clinicalTrials.gov*, mais ce n'est pas le seul. L'ensemble des registres d'essais cliniques est disponible à partir du **système d'enregistrement international des essais cliniques (ICTRP)** de l'Organisation mondiale de la santé, qui propose un moteur de recherche pour les consulter (www.who.int/ictrp/fr). L'interrogation de ces bases de données permet de mettre en évidence d'éventuelles études réalisées mais non publiées ou des études en cours de réalisation, ce qui permet de disposer des données les plus exhaustives possible [7].

Enfin, il est recommandé de consulter les références citées par les articles inclus dans la revue systématique, et par tout autre article pertinent. Inversement, il est possible de consulter l'ensemble des articles citant les études incluses dans la revue systématique, en utilisant par exemple **Web of Science** [8] (*Science Citation Index Expanded* et *Social Science Citation Index*), géré par le groupe Thomson Reuters.

Stratégie de recherche

La stratégie de recherche employée doit permettre une **recherche exhaustive**, afin de recueillir le plus grand nombre d'études et de diminuer les biais de sélection. Il n'existe pas de stratégie idéale, et chaque question doit conduire à une recherche adaptée. La stratégie de recherche doit reposer sur une liste de mots-clés et de synonymes, que l'on peut répertorier dans des articles de référence. Ces mots-clés, issus de texte libre ou de vocabulaire contrôlé, doivent figurer dans la recherche selon une logique précise. La méthode PICOT peut aider à l'identification d'une stratégie de recherche appropriée, en orientant l'utilisation des

mots-clés et des opérateurs booléens (AND, OR, NOT). Enfin, il est possible d'utiliser des stratégies de recherche prédéfinies pour filtrer les études en fonction de leur type, comme celles proposées par le groupe Cochrane pour identifier les essais contrôlés randomisés, dont un exemple est présenté en figure 1. La stratégie de recherche des études doit être élaborée lors de la mise en place du protocole et ne doit pas changer par la suite, il est donc indispensable qu'elle soit validée par tous les auteurs de la revue. La stratégie employée et les résultats obtenus doivent être documentés à chaque étape du processus de recherche.

```
#1 randomized controlled trial [pt]
#2 controlled clinical trial [pt]
#3 randomized [tiab]
#4 placebo [tiab]
#5 clinical trials as topic [mesh: noexp]
#6 randomly [tiab]
#7 trial [ti]
#8 #1 or #2 or #3 or #4 or #5 or #6 or #7
#9 animals [mh] not (humans [mh] and animals [mh])
#10 #8 not #9
```

Figure 1 : Stratégie de recherche de haute sensibilité pour identifier les essais randomisés dans MEDLINE

Version 2008, sensibilité et précision optimisées ; format PubMed. A partir du *Cochrane Handbook for Systematic Reviews of Interventions* [1], chapitre 6.4.11 « Search filters ».

Sélection des études

Lorsque la recherche des articles est terminée, l'importation des références dans un logiciel bibliographique tel qu'*EndNote*, *Zotero* ou *Reference Manager* peut permettre de trier les articles et d'éliminer les doublons. Il est ensuite nécessaire d'examiner les titres et/ou les résumés des études pour éliminer celles qui ne correspondent pas à la question de recherche. Les textes intégraux des études pertinentes doivent ensuite être lus afin d'évaluer leur éligibilité, en fonction des critères préalablement définis. Il peut être nécessaire de correspondre avec les investigateurs pour clarifier certains points ou récupérer des données manquantes. Il est utile de conserver une **liste des études exclues** et de préciser la raison de leur exclusion afin de montrer aux lecteurs que ces études ont bien été examinées mais non retenues. Cette liste doit contenir les études qui semblent *a priori* pertinentes, mais qui ne répondent pas aux critères d'éligibilité suite à une analyse approfondie. La sélection des études doit être faite par deux personnes de façon indépendante, avec un mécanisme de conciliation

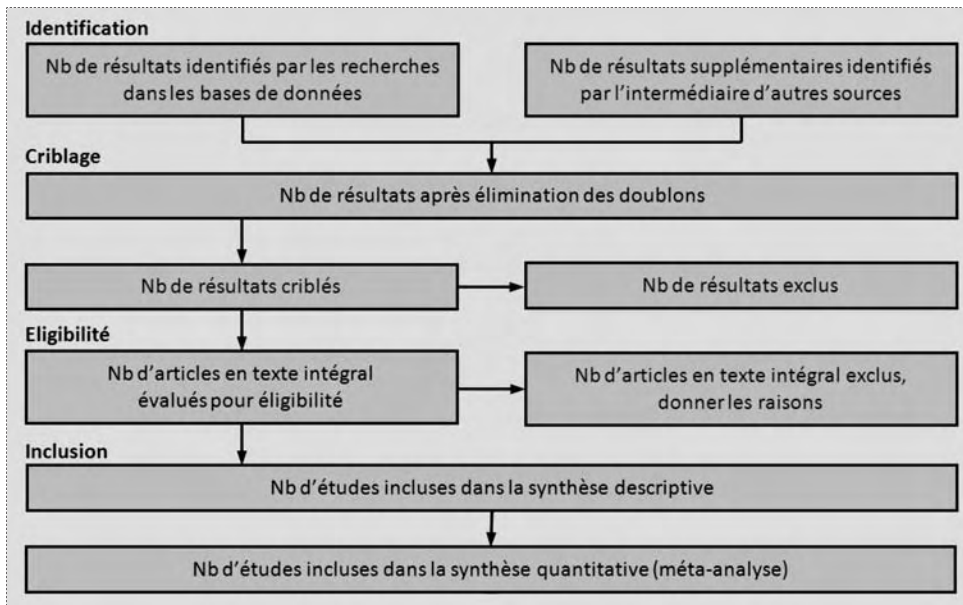


Figure 2 : Exemple de *flow chart* (ou diagramme de flux) illustrant le processus de sélection des études

Traduction libre à partir de Moher *et al.*, *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement* [9].

S.F.S.P. | Téléchargé le 06/06/2026 sur <https://strm.cairn.info> (IP: 216.73.217.8)

pour la résolution des désaccords. La construction d'un **diagramme de flux**, ou *flow chart*, permet d'illustrer le processus de recherche et de sélection des études. Un modèle de diagramme de flux est proposé dans un guide pour la rédaction des revues systématiques, réalisé par Moher *et al.* [9] (figure 2).

Extraction des données

Après la décision finale d'inclusion des études dans la revue, il est procédé à l'extraction des données. Celles-ci doivent être extraites des études par au moins deux personnes indépendantes. L'utilisation d'un **formulaire de collecte de données** permet de recueillir l'ensemble des données nécessaires à l'analyse des études, comme les critères d'éligibilité, les méthodes, les participants, les interventions, les critères de jugement ou les résultats. Les risques de biais doivent également être extraits afin d'évaluer le niveau de preuve de chaque étude. L'utilisation d'un formulaire de collecte de données est indispensable afin de permettre par la suite d'effectuer les analyses nécessaires avec une bonne qualité méthodologique. Il existe des formulaires types, comme le *Cochrane data collection form for intervention review* [2], qui doivent être préalablement testés afin d'être éventuellement adaptés aux objectifs de la revue. Lorsque toutes les données ont été

extraites et recueillies dans le formulaire de collecte, l'utilisation d'un logiciel de gestion de données comme Microsoft Excel ou **RevMan**, qui est conçu pour la réalisation de revues systématiques Cochrane, permet de préparer l'analyse des données.

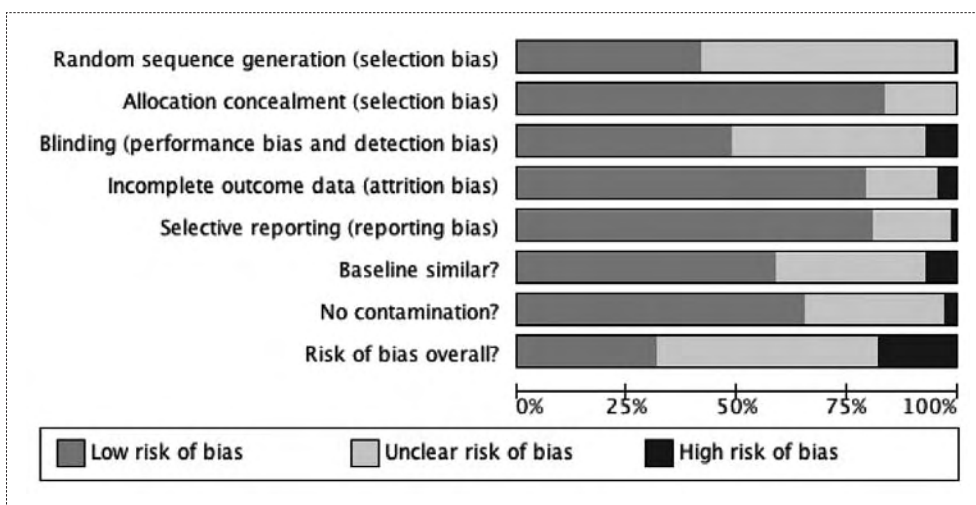
Évaluation de la qualité méthodologique des études

Une évaluation de la qualité méthodologique des études doit systématiquement accompagner l'analyse des résultats. Celle-ci est une composante essentielle de la revue systématique, et influence largement ses conclusions. S'ils ne sont pas détectés, des problèmes de conception ou de réalisation des études peuvent générer des interprétations erronées et fausser les résultats d'une revue. Ainsi, l'évaluation des études incluses dans une revue systématique doit toujours comporter une **évaluation des risques de biais**. Un biais est une cause d'erreur systématique pouvant conduire à une sous-estimation ou une surestimation du réel effet d'une intervention. De nombreux outils permettent d'évaluer la qualité méthodologique des études d'interventions. Le groupe EPOC de la collaboration Cochrane a par exemple développé une liste de risques de biais pour les études d'interventions visant à améliorer la qualité des soins. Les risques de biais à évaluer sont différents selon le type

Tableau II : Suggestion des risques de biais pour les revues EPOC. Risques de biais concernant les études avec groupe contrôle (essais randomisés et études avant-après). Traduction libre à partir de *Suggested risk of bias criteria for EPOC reviews*, du groupe EPOC [2]

Le processus d'allocation était-il correct ?	Faible risque : si un processus de randomisation est décrit. Haut risque : si une méthode non aléatoire est utilisée. Les essais avant-après sont à haut risque. Risque incertain : si non spécifié.
L'allocation était-elle correctement masquée ?	Faible risque : si l'unité d'allocation est un professionnel, une équipe ou une institution, toutes les unités doivent être réparties au début de l'étude. Si l'unité d'allocation est le patient ou un épisode de soin, la randomisation doit être masquée par une méthode appropriée. Haut risque : si la répartition des sujets n'est pas dissimulée. Les essais avant-après sont à haut risque. Risque incertain : si non spécifié.
Les niveaux de base des critères de jugement étaient-ils similaires ?	Faible risque : si les niveaux de base des critères de jugement sont mesurés avant l'intervention et qu'il n'y a pas de différence importante entre les groupes, ou si les différences sont ajustées pour l'analyse. Haut risque : si des différences importantes entre les groupes existent et ne sont pas ajustées pour l'analyse. Risque incertain : si les niveaux de bases des critères de jugement ne sont pas mesurés.
Les caractéristiques de base étaient-elles similaires ?	Faible risque : si les caractéristiques de base des professionnels de santé sont reportées et similaires. Haut risque : si les caractéristiques de base des professionnels de santé ne sont pas mentionnées ou s'il y a des différences entre les professionnels de santé des groupes contrôle et intervention. Risque incertain : si les caractéristiques sont mentionnées mais les données ne sont pas reportées.
Les données incomplètes ont-elles été correctement traitées ?	Faible risque : s'il est peu vraisemblable que les données manquantes faussent les résultats (proportions similaires dans les groupe intervention et contrôle, ou faible proportion de données manquantes). Haut risque : si les données manquantes peuvent fausser les résultats. Risque incertain : si non spécifié. Ne pas présumer 100 % de suivi si ce n'est pas explicité.
La connaissance de la répartition des sujets dans les groupes a-t-elle été correctement prévenue ?	Faible risque : si les critères de jugement principaux ont été mesurés en aveugle, ou si les critères étaient objectifs. Haut risque : si les critères de jugement principaux n'ont pas été mesurés en aveugle et que cela peut biaiser les résultats. Risque incertain : si non spécifié.
Le risque de contamination a-t-il été correctement prévenu ?	Faible risque : si l'unité d'allocation était une communauté, une institution ou un cabinet et qu'il est peu vraisemblable que le groupe contrôle ait reçu l'intervention. Haut risque : s'il est vraisemblable que le groupe contrôle ait reçu l'intervention, c'est-à-dire si les patients plutôt que les professionnels ont été randomisés. Risque incertain : si les professionnels ont été randomisés au sein d'un hôpital ou d'un cabinet, et que les professionnels des groupes contrôle et intervention ont pu communiquer entre eux.
L'étude est-elle exempte d'un report sélectif des critères de jugement ?	Faible risque : si l'élément faisant suspecter un report sélectif des critères de jugement, c'est-à-dire si tous les critères de jugement présents dans la partie « méthodes » sont reportés dans la partie « résultats ». Haut risque : si des critères de jugement importants ne sont pas présentés dans la partie « résultats ». Risque incertain : si non spécifié.

Figure 3 : Exemple de représentation graphique des risques de biais, exprimés en pourcentage des études incluses dans la revue systématique



de l'étude. Par exemple, pour un essai contrôlé randomisé, huit principaux points doivent être considérés, concernant notamment la randomisation, l'aveugle, la comparabilité des groupes, le traitement des données incomplètes ou encore la prévention de la contamination (tableau II). Pour chaque item, les auteurs doivent juger si le risque est correctement prévenu ou non (faible risque de biais, haut risque ou risque incertain) en justifiant leur choix. Afin de s'assurer que les risques de biais sont correctement jugés, **au moins deux personnes** doivent procéder de façon indépendante à leur évaluation. Les risques de biais peuvent ensuite être synthétisés sous forme graphique, comme l'illustre la figure 3. Pour une même taille d'effet, il est évident que le **niveau de preuve** sera différent si la plupart des études sont à faible risque ou à risque élevé de biais.

Synthèse des données

Le mode de synthèse des données (types de comparaisons, critères de jugement, méthode statistique utilisée) doit avoir été précisé dans le protocole initial de la revue. Selon les études incluses dans la revue, les données peuvent être synthétisées de manière quantitative, semi-quantitative ou descriptive. L'**analyse quantitative**, combinaison statistique des résultats d'au moins deux études distinctes, le plus souvent réalisée par la technique de la **méta-analyse**, n'est possible qu'à partir d'études relativement homogènes. Elle n'est donc pas toujours appropriée, par exemple lorsque les bras traitements ou comparateurs sont très différents entre les études, lorsque les critères de jugement ne sont pas homogènes, ou lorsque les études présentent des risques de biais élevés. C'est particulièrement le cas dans le domaine des pratiques et de l'organisation des soins, où les interventions, les systèmes de soins et les populations peuvent être trop hétérogènes pour permettre une synthèse quantitative. Dans ce type de cas, il est parfois plus approprié de réaliser une **synthèse « semi-quantitative »**, ou de ne présenter qu'une **synthèse descriptive**.

Synthèse quantitative par méta-analyse

Calcul des indices d'effet

Les données brutes des études doivent être ré-analysées : pour chaque étude, l'effet de l'intervention doit être quantifié grâce à une mesure appropriée, l'**indice d'effet**. Cet

indice indique l'ampleur et la direction d'une association entre des variables, et intègre la notion de variabilité du résultat sous forme d'intervalle de confiance. L'indice d'effet s'exprime de manière différente en fonction du type de données collectées. Ainsi dans le cas de variables dichotomiques, on utilisera le plus souvent le **risque relatif** ou l'**odds ratio**, alors que pour des variables continues on utilisera généralement une **différence de moyennes**, c'est-à-dire la différence entre le résultat moyen du groupe intervention et le résultat moyen du groupe contrôle. Dans ce cas, si les résultats ne sont pas exprimés dans la même unité dans toutes les études, il est nécessaire de calculer une **différence de moyennes standardisée**, ou effet standardisé (on retrouve parfois le terme *effect size*, mais la collaboration Cochrane recommande de ne plus l'utiliser à cet effet). Pour cela, on rapporte la différence de moyennes de chaque étude à l'**écart-type** du résultat moyen de l'étude. La validité de ce mode d'expression repose sur l'hypothèse que la variabilité est la même dans toutes les populations étudiées, et que les différences entre les écarts-types ne sont dues qu'aux différences d'unité de mesure. Cela peut poser problème, par exemple si une population est plus hétérogène que les autres : dans ce cas son écart-type risque d'être plus élevé, et donc la différence de moyennes standardisée sera plus proche de zéro. Lorsque tous les indices d'effet ont été obtenus, ceux-ci doivent ensuite être combinés de manière statistique pour déterminer un **effet commun**.

Agrégation des données

L'estimation de l'**effet commun** et de sa variance pour l'ensemble des études incluses s'effectue en combinant les effets de chaque étude, pondérés en fonction de leur précision ou de la taille de leur échantillon. Ainsi, plus les études ont de grands échantillons et un intervalle de confiance étroit, plus leur contribution à l'effet commun est importante. Il est à noter que le calcul de l'effet commun n'est juste, d'un point de vue statistique, que si les indices d'effet sont indépendants les uns des autres, ce qui implique que les échantillons des études incluses soient indépendants. Ceci exclut par exemple les études dans lesquelles deux groupes subissant une intervention différente sont comparés à un même groupe contrôle. Dans ce cas il n'est pas possible de calculer un effet commun qui intégrerait les deux interventions, car la population du groupe contrôle serait utilisée pour calculer deux tailles d'effet.

Lorsque les indices d'effet pondérés ont été calculés, la réalisation d'un **test d'hétérogénéité** permet d'établir si les différences entre les effets des interventions peuvent

être attribuées à l'« erreur d'échantillonnage », c'est-à-dire au hasard, ou s'il existe des différences réelles entre les effets des interventions. L'hétérogénéité peut être appréciée visuellement sur la représentation graphique, ou par des tests statistiques tels que le χ^2 , dont la puissance de détection est assez faible, ou le calcul du « I^2 ». Ce dernier représente une estimation du pourcentage de la variabilité due à l'hétérogénéité [10]. Une valeur de I^2 inférieure à 25 % indique une hétérogénéité faible, une valeur comprise entre 25 % et 50 % une hétérogénéité modérée, et une valeur supérieure à 50 % une hétérogénéité importante. L'hétérogénéité est inhérente à la méta-analyse et ne doit pas être considérée comme un problème devant être éliminé, mais au contraire comme un paramètre très informatif pour l'extrapolation des résultats [11, 12]. Son exploration par des techniques appropriées est indispensable au processus de méta-analyse. Il faut rechercher les composantes des études susceptibles de modifier l'effet de l'intervention. Par exemple, la durée de l'intervention, le type de professionnels ciblés, ou encore l'état des pratiques avant intervention peuvent être responsables de l'hétérogénéité observée.

Les résultats du test d'hétérogénéité permettent d'orienter le choix du modèle d'analyse à privilégier pour estimer l'effet commun : deux modèles peuvent être utilisés, se distinguant essentiellement par la définition qu'ils font de la variation entre les échantillons. Dans les **modèles à effets fixes**, utilisables quand le test d'hétérogénéité est négatif, on considère que toutes les études proviennent d'une même population dans laquelle l'intervention produit une seule taille d'effet. La variation inter-échantillons est donc exclusivement attribuée à l'erreur d'échantillonnage. À l'opposé, les **modèles à effets aléatoires** sont utilisés lorsqu'il existe une hétérogénéité entre les effets des interventions. Dans ce type de modèle, on postule que les différences entre les effets des interventions s'expliquent non seulement par le hasard, mais aussi par des variations réelles entre les études. Lorsque l'hétérogénéité entre les études est très importante, la question de la pertinence d'une synthèse quantitative se pose. En effet, est-il raisonnable de calculer un effet commun quand l'effet de l'intervention est extrêmement variable selon les interventions ? Dans ce cas, il est parfois préférable de se contenter d'une synthèse « semi-quantitative » ou descriptive.

Lorsque l'effet commun a été calculé, un **test d'association** est réalisé. Son seuil de signification est généralement fixé à 5 %. Quand il est significatif, on considère qu'il existe un effet statistiquement significatif de l'intervention. S'il n'est pas significatif, se pose la question de la puissance de la méta-analyse, qui peut être calculée *a posteriori*.

Analyse de sensibilité

Le processus de réalisation d'une revue systématique implique une succession de prises de décisions, parfois arbitraires ou subjectives. Par exemple, la décision d'inclure dans la méta-analyse des essais autres que randomisés, le choix d'une tranche d'âge ou d'un degré de sévérité d'une maladie dépendent en grande partie de la volonté des auteurs. L'**analyse de sensibilité** permet de vérifier que les résultats et les conclusions de la méta-analyse sont suffisamment robustes et ne sont pas affectés par des décisions discutables prises lors du processus d'analyse. Elle consiste à vérifier que les résultats restent les mêmes lorsqu'on répète l'analyse en remplaçant ces décisions par d'autres. Par exemple, l'analyse peut être effectuée en ne prenant en compte que les essais de meilleure qualité méthodologique, en excluant les essais avec des données manquantes, ou en se limitant à des sous-groupes particulièrement importants ou à certains types d'interventions. Lorsque l'analyse de sensibilité est négative, on peut considérer que les décisions prises lors du processus d'analyse n'ont pas influencé les résultats. Dans le cas contraire, il peut être nécessaire de résoudre certaines ambiguïtés, de réclamer des informations supplémentaires auprès des auteurs, ou de n'inclure dans la méta-analyse que certaines études, telles que les plus précises ou celles de meilleure qualité méthodologique par exemple. Les conclusions de la revue systématique doivent alors être formulées avec précaution, d'autant plus que cette problématique méthodologique est encore largement discutée [13]. Certaines analyses de sensibilité peuvent être prévues dans le protocole de la revue systématique, mais beaucoup ne se révèlent nécessaires qu'au moment du processus d'analyse des données.

Expression des résultats

Les résultats de la méta-analyse sont généralement représentés sous forme graphique, ce qui permet d'apprécier visuellement l'effet commun de l'intervention. Le type de graphique le plus utilisé est le **forest plot** (Figure 4), dans lequel une ligne verticale correspond à l'absence d'effet de l'intervention, c'est-à-dire la valeur 1 pour un risque relatif ou un odds ratio, ou la valeur 0 pour une différence de moyenne. L'indice d'effet de chaque étude est symbolisé par un rectangle et une ligne horizontale, qui représentent respectivement l'estimation de l'effet et son intervalle de confiance. Si l'intervalle de confiance coupe la ligne verticale, l'effet de l'intervention est considéré comme non significatif. L'aire du rectangle est proportionnelle à la

pondération de l'étude, c'est-à-dire au poids de l'étude dans l'analyse statistique, qui dépend de sa précision ou de la taille de sa population. Un losange représente la combinaison pondérée des résultats individuels (effet commun). Si celui-ci ne chevauche pas la ligne verticale, on considère que l'intervention a un effet significatif.

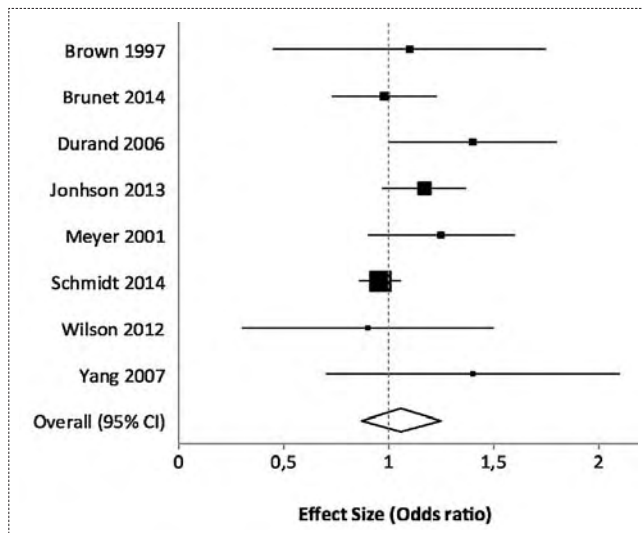


Figure 4 : Exemple fictif de *forest plot*

Dans cet exemple, l'effet commun des interventions n'est pas significatif, car le losange chevauche la ligne pointillée verticale représentant l'absence d'effet. L'étude ayant le plus de poids dans la méta-analyse est celle de Schmidt.

Un autre type de graphique peut être retrouvé, le *funnel plot*, ou graphique en entonnoir (Figure 5). Celui-ci présente les résultats des études individuelles de manière à détecter un éventuel biais de publication des études [14]. Dans ce graphique, l'ordonnée correspond à la précision de l'étude (et donc à la taille de sa population), et l'abscisse correspond à l'effet de l'intervention. L'effet commun de l'intervention est représenté par une ligne verticale continue. Plus une étude est précise, plus son intervalle de confiance est étroit, et plus haut elle est placée dans le *funnel plot*. Les résultats des études les plus précises sont généralement proches de l'effet commun, et au contraire les résultats des études moins précises ont plus de chance d'en être éloignés, car ils sont davantage affectés par le hasard. Si ces hypothèses sont respectées, le *funnel plot* prend une forme symétrique et pyramidale. Des anomalies dans l'aspect du *funnel plot* peuvent suggérer un biais de publication. Ainsi, dans cet exemple, on constate que les études les plus précises sont à gauche de l'effet commun, et que les études les moins précises ont tendance à être placées à droite. On

peut supposer que les petites études ne montrant pas l'effet escompté ne sont pas toujours publiées, à la différence des grosses études qui le sont plus systématiquement.

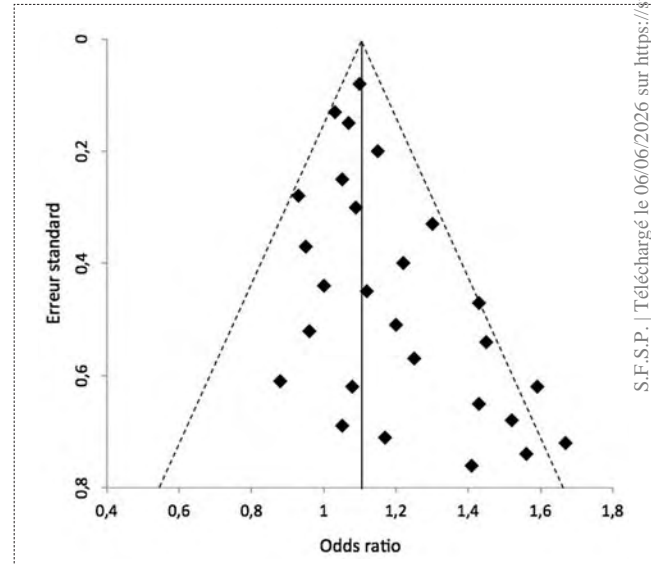


Figure 5 : Exemple fictif de *funnel plot*

Cet exemple de *funnel plot* traduit un risque de biais de publication : les études les moins précises (situées en bas) ne sont pas réparties de manière symétrique autour de l'effet commun.

Synthèse dite « semi-quantitative »

Lorsque les populations, les interventions ou les critères de jugement sont trop hétérogènes, la réalisation d'une méta-analyse peut être difficile. Dans ce cas, il est possible d'évaluer l'effet des interventions d'une manière que nous appellerons « **semi-quantitative** ». Ce type d'analyse est utilisé dans de nombreuses revues systématiques du groupe EPOC consacrées aux interventions dites « professionnelles », destinées à améliorer les pratiques cliniques (formation médicale continue, diffusion de l'information, leaders d'opinion ou encore rappels au moment de la décision).

L'analyse principale est centrée sur l'analyse des variables dichotomiques : la pratique est jugée appropriée ou non appropriée, permettant de calculer un pourcentage d'adéquation aux recommandations. Pour chaque étude, lorsque plusieurs critères de jugement sont présentés, c'est le critère de jugement principal de l'étude (lorsqu'il est défini par les auteurs de l'article) qui est pris en compte. Si le critère de jugement principal n'est pas défini et que plusieurs variables dichotomiques sont présentées, la valeur médiane de ces différentes variables est utilisée. Le résultat de la revue systématique est présenté sous la forme d'un effet médian,

amélioration absolue médiane du pourcentage d'adéquation des pratiques dans le groupe intervention comparée au groupe contrôle.

Dans le domaine des pratiques professionnelles et de l'organisation des soins, la synthèse doit prendre en compte certaines caractéristiques spécifiques de ce type d'études. Par exemple, l'existence ou non d'une analyse des pratiques avant intervention, ou le type de schéma expérimental utilisé. Ces caractéristiques ont été développées dans un article précédent de la revue [15]. Que la synthèse soit quantitative ou semi-quantitative, il est recommandé de présenter les résultats par type d'étude [16]. Ainsi, les résultats des essais randomisés doivent être présentés séparément des études quasi expérimentales, c'est-à-dire séries temporelles interrompues et essais contrôlés avant et après.

Synthèse descriptive

Elle correspond à une présentation structurée [17] des caractéristiques des études, qui commence par une **analyse descriptive** de chaque étude, en ce qui concerne sa méthodologie, l'intervention évaluée ou encore ses résultats. Les études doivent ensuite être associées en **catégories logiques**, ce qui permet d'orienter les conclusions de la revue systématique : un regroupement des études en fonction du type d'intervention ou des résultats permet de déterminer quel type d'intervention semble le plus efficace, un regroupement par population ou contexte permet de juger le cadre qui paraît le plus favorable à l'intervention, un regroupement par méthodologie permet de rassembler les études les plus probantes. Après que les études aient été analysées et regroupées, la **synthèse de leurs résultats** permet de produire un sommaire des résultats, en tenant compte des différences entre les études pouvant influencer leurs conclusions. Les auteurs donnent parfois un résultat chiffré correspondant au pourcentage d'études « positives » (c'est-à-dire les études dont plus de 50% des critères de jugement sont significativement améliorés). Malgré l'absence de résultat quantitatif, les synthèses descriptives sont des documents utiles permettant de faire le point sur un sujet donné.

Analyse de l'aspect qualitatif d'une intervention

Si l'on sait bien évaluer la quantité d'effet d'une intervention, il est également intéressant de savoir pourquoi et comment celle-ci fonctionne. On peut pour cela se pencher sur l'aspect « qualitatif » de l'intervention, évalué dans certaines études sous la forme de l'opinion des participants ou d'aspects socio-comportementaux. Ces **données « qualitatives »** permettent de comprendre

comment l'intervention est vécue par les patients ou les professionnels. Elles peuvent être prises en compte de manière complémentaire à l'analyse de l'effet de l'intervention lors de la réalisation d'une revue systématique. Ce type d'approche enrichit la réflexion autour des interventions et de leur efficacité, en permettant par exemple de comprendre pourquoi certaines interventions sont efficaces et d'autres non. Par exemple, il a été montré que la relation patient-médecin et les croyances des patients pouvaient limiter l'effet d'interventions visant à réduire la prescription d'antibiotiques [18]. Ivers *et al.* [19] ont, quant à eux, mis en évidence qu'un *feedback* aux professionnels de santé était plus efficace sur les pratiques quand il était donné par un supérieur ou un collègue reconnu. Des interventions qui semblent identiques peuvent ainsi avoir une efficacité différente en fonction de certains facteurs qu'il est important de détecter et de prendre en compte. L'intégration de données qualitatives dans une revue systématique est cependant un défi méthodologique, en raison de leur nature subjective et hétérogène. Ces données peuvent être traitées de manière descriptive ou agrégative, en veillant à la rigueur et à la transparence du processus. Il existe de nombreuses méthodes de synthèse des données qualitatives faisant l'objet de discussions au sein de la collaboration Cochrane et du « **Cochrane Qualitative and Implementation Methods Group** » (CQIMG) [20].

Synthèse du niveau de preuve

Une revue systématique permet donc de répondre à une question posée : l'intervention a-t-elle un effet, et quelle est son importance ? Mais quel niveau de confiance peut-on accorder au résultat ? En d'autres termes, quel est le niveau de preuve des conclusions de la revue ? Cela dépend de différents facteurs comme le type d'études, les risques de biais, l'hétérogénéité des effets, les types de comparaisons effectuées dans les études ou encore l'imprécision des résultats. Le groupe **GRADE** (*The Grades of recommendation, assessment, development and evaluation working group*) [21-23] propose une approche intéressante, de plus en plus utilisée dans les revues systématiques de la collaboration Cochrane, permettant d'attribuer pour chaque critère de jugement un **niveau de preuve** « élevé », « modéré », « faible » ou « très faible ». Le groupe GRADE a publié de nombreux articles et documents pédagogiques permettant au lecteur intéressé de se familiariser avec cette méthode (www.gradeworkinggroup.org).

Lorsqu'une revue systématique est utilisée pour formuler une recommandation, le niveau de preuve de ses résultats est l'un des éléments permettant de déterminer la **force de la recommandation**, au même titre que le rapport bénéfices-inconvénients de l'intervention, les préférences ou encore le contexte économique. En fonction de ces éléments, le système GRADE attribue à une intervention l'une des deux catégories de recommandations : forte ou faible. La Haute Autorité de santé a réalisé un état des lieux des différents systèmes de définition des niveaux de preuve et de gradation des recommandations [24].

Rédaction et publication de l'article

L'écriture d'une revue systématique doit respecter une méthodologie rigoureuse afin de garantir sa qualité et de permettre son évaluation par les lecteurs. Pour cela, les auteurs ont à leur disposition des guides permettant de standardiser le format des rapports de revues systématiques. Développé par Moher *et al.* en 2009 [9], le **PRISMA statement** (*Preferred Reporting Items for Systematic reviews and*

Tableau III : Version française de la grille AMSTAR, adaptée du Centre de collaboration nationale des méthodes et outils (CCNMO)

	A-t-on fourni un protocole <i>a priori</i> du travail ?
1	La question à l'étude et les critères d'inclusion doivent être établis avant l'exécution de la revue systématique. Existe-t-il un protocole publié de l'étude, l'approbation par un comité d'éthique ou des objectifs prédéterminés ?
	La sélection des études et l'extraction des données ont-elles été faites en double ?
2	Il doit y avoir au moins deux extracteurs de données indépendants et un mécanisme pour arriver à un consensus dans les cas de divergences.
	A-t-on effectué une recherche complète dans la littérature ?
3	La recherche doit porter sur au moins deux sources électroniques. Le rapport doit inclure les années et les bases de données utilisées (ex. : Central, EMBASE et MEDLINE). Les auteurs doivent fournir les mots clés et/ou les termes utilisés pour la recherche et au mieux la stratégie de recherche. Toutes les recherches doivent être complétées par une consultation de revues, de manuels, de registres spécialisés ou de spécialistes du domaine et par une revue des références contenues dans les études.
	Est-ce que la nature d'une publication (ex. : littérature grise) a servi de critère d'inclusion ?
4	Les auteurs doivent déclarer avoir recherché des rapports d'études sans égard au type de publication. Ils doivent aussi dire s'ils ont exclu des rapports à cause de la nature de la publication, de sa langue, etc. Il doit y avoir une recherche explicite de la littérature « grise », comme par exemple la base OpenGrey.
	Les auteurs ont-ils fourni la liste des études incluses et des études exclues ?
5	Les auteurs doivent fournir la liste des études incluses et des études exclues. Un lien électronique vers ces listes est acceptable.
	Les auteurs ont-ils fourni une description des caractéristiques des études incluses ?
6	Présentées sous une forme condensée comme un tableau, les données de l'étude originale doivent inclure les participants, les interventions et les résultats. L'étude doit rendre compte des différentes caractéristiques de toutes les études analysées (ex. : âge, sexe, données socioéconomiques pertinentes, état de la maladie, durée, sévérité ou autres maladies).
	La qualité scientifique des études incluses a-t-elle été analysée et documentée ?
7	Le choix des types d'études éligibles doit être précisé et justifié (par exemple les essais contrôlés randomisés, les essais en double insu, les essais comparatifs avec placebo, d'autres types d'études). La qualité de chaque étude a-t-elle été analysée ? (score de Jadad, risque de biais...).
	La qualité des études incluses a-t-elle été prise en compte de façon appropriée dans la formulation des conclusions ?
8	Les résultats concernant la qualité scientifique des études incluses doivent être pris en compte dans l'analyse et les conclusions de la revue, et doivent être mentionnés explicitement dans la formulation des recommandations. Par exemple « les résultats doivent être interprétés avec précaution en raison de la faible qualité des études prises en compte ».
	Les méthodes d'analyse des résultats étaient-elles appropriées ?
9	Lorsqu'on regroupe des résultats, on doit d'abord vérifier si les études sont combinables en appliquant un test d'homogénéité (p. ex., le test I ²). S'il existe de l'hétérogénéité, on devrait utiliser un modèle d'analyse des effets aléatoires et/ou considérer la pertinence de regrouper les résultats des études (est-il approprié de les regrouper ?). Les auteurs doivent expliquer la raison de leur mode de présentation des résultats.
	A-t-on analysé la possibilité d'un biais de publication ?
10	L'analyse d'un biais de publication doit comprendre une représentation graphique (p. ex., graphique en entonnoir) et/ou des analyses statistiques (p. ex. : le test de régression de Egger).
	A-t-on déclaré les conflits d'intérêts ?
11	Les sources potentielles de soutien doivent être clairement présentées, pour les auteurs de la revue systématique et pour chacune des études incluses dans la revue.

Meta-Analyses) est une mise à jour du QUOROM statement, qui ne devrait plus être utilisé. Le PRISMA statement est constitué d'une check-list de 27 points et d'un *flow chart* (figure 2), permettant de faciliter l'écriture tout autant que la lecture de la revue systématique. La collaboration Cochrane a également élaboré son propre guide, **MECIR** (*Methodological Expectations of Cochrane Intervention reviews*).

Il est également souhaitable de définir préalablement qui seront les auteurs, et quelles seront leurs contributions et leurs positions dans la signature de l'article. Chaque auteur doit mentionner ses éventuels conflits d'intérêt, ainsi que ceux des études incluses, afin d'assurer la transparence de la revue systématique. D'autres recommandations relatives à l'écriture et à la publication des articles scientifiques sont proposées par l'**International Committee of Medical Journal Editors** (ICMJE), ou convention de Vancouver [25]. Il est bien entendu essentiel de tenir compte des recommandations aux auteurs spécifiques à chaque journal. Certains journaux sont spécialisés dans la publication de revues systématiques, comme le *systematic reviews journal*, diffusé par BioMed Central. Ce journal publie également des articles méthodologiques.

améliorer la qualité des soins. Une série d'articles intitulée « *increasing value and reducing waste* » [29] publiée dans *The Lancet* recommande que toute nouvelle étude soit justifiée par une revue systématique résumant l'état des connaissances, et que les décideurs et les organismes de financement orientent davantage la recherche vers des synthèses de données existantes. La réalisation d'une revue systématique s'inscrit dans le concept de **médecine factuelle**, ou **evidence-based medicine**, défini par Sackett *et al.* [30] comme « l'utilisation consciencieuse, explicite et judicieuse des meilleures données actuelles pour la prise de décision dans les soins à prodiguer aux patients individuels ». Dans le domaine de l'organisation des soins, comme les populations et les organisations sont relativement différentes d'un système de santé à un autre, il est assez difficile de conclure de manière formelle et générale sur l'effet d'une intervention. La revue systématique fournit des éléments permettant de guider les pratiques, à condition de tenir compte des caractéristiques du système de soins auquel on veut appliquer l'intervention [31].

Aucun conflit d'intérêt déclaré

Évaluation des revues systématiques publiées

Après sa publication, la qualité méthodologique d'une revue systématique peut être analysée grâce à un outil de référence, la **grille AMSTAR**. Cet outil, mis au point en 2007 par Shea *et al.* [26], repose sur 11 questions portant notamment sur le processus de sélection des études, les caractéristiques des études retenues, l'évaluation de la qualité des études ou encore sur la méthodologie employée pour l'analyse statistique (tableau III). En utilisant cet outil, de nombreux auteurs ont montré que la qualité méthodologique des revues systématiques publiées n'était souvent pas optimale [27, 28]. Il est pourtant indispensable de respecter une méthodologie rigoureuse afin d'obtenir des résultats de qualité, qui puissent avoir un impact sur les pratiques ou l'organisation des soins.

Conclusion

Dans le contexte d'augmentation exponentielle de l'information médicale, la revue systématique est un outil essentiel pour évaluer l'efficacité des interventions visant à

Références

- Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2008. Available from <www.cochrane-handbook.org>.
- Norwegian Satellite of the Cochrane Effective Practice and Organisation of Care Group [Internet]. Oslo ; [Cited: 2014 September 10] Available from: <http://epocoslo.cochrane.org>.
- Booth A, Clarke M, Dooley G, Ghera D, Moher D, Petticrew M, *et al.* The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev.* 2012;1:2.
- Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev.* 2007;2.
- OpenGrey [Internet]. [cited: 2014 September 10] Available from: <http://www.opengrey.eu>.
- Grey Literature Report [Internet]. [cited 2014 Mar 6] Available from: <http://www.greylit.org>.
- Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med.* 2013;10(12):e1001566.
- Web of Science - Thomson Reuters [Internet]. [cited 2014 September 10] Available from: <http://wokinfo.com>.
- Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535-b2535.

10. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-60.
11. Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? *Stat Med*. 1987; 6(3):351-60.
12. Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. *JAMA*. 2014;312(6):603-5.
13. Dechartres A, Altman DG, Trinquart L, Boutron I, Ravaud P. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. *JAMA*. 2014;312(6):623-30.
14. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997; 315(7109):629-34.
15. Zaugg V, Savoldelli V, Sabatier B, Durieux P. Améliorer les pratiques et l'organisation des soins : méthodologie des études d'interventions. *Santé Publique*. 2014;26(4):519-30.
16. Egger M, Smith GD, Altman D. *Systematic reviews in health care: meta-analysis in context*. John Wiley & Sons; 2008.
17. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, *et al*. Guidance on the conduct of narrative synthesis in systematic reviews. *Prod ESRC Methods Programme Version* ; 2006.
18. Arroll B, Goodyear-Smith F, Thomas DR, Kerse N. Delayed antibiotic prescriptions: what are the experiences and attitudes of physicians and patients? *J Fam Pract*. 2002;51(11):954-9.
19. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;6.
20. Cochrane Qualitative and Implementation Methods Group [Internet]. [cited 2014 September 10] Available from: <<http://cqim.cochrane.org/welcome>>.
21. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, *et al*. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
22. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-6.
23. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A, *et al*. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol*. 2011;64(4):380-2.
24. Haute Autorité de Santé. Niveau de preuve et gradation des recommandations de bonne pratique. 2013.
25. ICMJE [Internet]. [cited 2014 September 10] Available from: <<http://www.icmje.org>>
26. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, *et al*. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
27. Fleming PS, Koletsi D, Seehra J, Pandis N. Systematic reviews published in higher impact clinical journals were of higher quality. *J Clin Epidemiol*. 2014;67(7):754-9.
28. Melchioris AC, Correr CJ, Venson R, Pontarolo R. An analysis of quality of systematic reviews on pharmacist health interventions. *Int J Clin Pharm*. 2012;34(1):32-42.
29. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, *et al*. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383(9912):156-65.
30. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71.
31. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet*. 2003; 362(9391):1225-30.