

Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
Ecole doctorale 393 Pierre Louis de santé publique
Spécialité : Epidémiologie et science de l'information

Centre de Recherche Épidémiologie et Statistiques Sorbonne Paris Cité (CRESS)
U1153, équipe METHODS : Méthodes en Évaluation Thérapeutique des Maladies
Chroniques

Le gaspillage évitable de la recherche dans les essais cliniques

Par Youri Yordanov

Dirigée par le Dr Agnes DECHARTRES et le Pr Philippe RAVAUD

Présentée et soutenue publiquement à L'Hôtel-Dieu, Paris le 20 Mars 2019

Président du jury : Professeur Bertrand Guidet, Sorbonne Universités

Rapporteur : Professeur Frédéric Adnet, Université Paris XIII

Rapporteur : Professeur Florence Canoui-Poitaine, Université Paris XII

Examineur : Professeur Catherine Paugam, Université Paris Diderot

Directeur de thèse : Professeur Philippe Ravaud, Université Paris Descartes

Co-directeur de thèse : Docteur Agnes Dechartres, Sorbonne Universités



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Remerciements

En tout premier lieu, je tiens à remercier et exprimer ma profonde gratitude à **Monsieur le Professeur Philippe Ravaud** qui a guidé mes premiers pas en recherche dès mon Master 2, m'a accueilli dans son équipe de recherche, a accepté de codiriger cette thèse et m'a fait bénéficier de son expertise et de ses conseils avisés.

J'accorde évidemment une pensée très particulière et mes plus sincères remerciements à **Madame la Docteur Agnès Dechartres**, qui a accepté de codiriger cette thèse. Je voudrais la remercier de sa patience et persévérance, d'avoir cru en moi et en mes capacités. Je suis très heureux qu'elle ait guidé mes premiers pas en recherche et j'espère que notre collaboration scientifique perdurera au-delà des travaux déjà menés.

Je souhaite exprimer toute ma gratitude envers les membres du jury qui ont accepté de juger mon travail. J'adresse mes sincères remerciements à **Madame la Professeur Florence Canouï-Poitrine** et **Monsieur le Professeur Frédéric Adnet** qui ont bien voulu être rapporteurs de ce travail, et à **Madame la Professeur Catherine Paugam** et **Monsieur le Professeur Bertrand Guidet** qui ont accepté de participer à ce jury de thèse.

Merci aussi...

Je souhaite remercier particulièrement Élise Diard sans l'imagination et créativité de laquelle les figures des articles inclus dans cette thèse n'auraient jamais vu le jour.

Ensuite, je tiens à remercier tous les membres du Centre d'Épidémiologie clinique, de l'unité INSERM U1153, pour les bons moments passés ensemble depuis quelques années ; en particulier Aïda, Alexandre, Ali, Amélie, Anthony, Arthur, Audrey, Aurélie, Aymar, Carolina, Céline, Isabelle, Ignacio, re Isabelle, Medhi, Mickael, Perrine, Raphaël, Stéphanie, Romana et bien évidemment l'inénarrable et inépuisable TTV.

Encore des remerciements, mais cette fois à la joyeuse communauté des urgences et en particulier des urgences de l'Hôtel-Dieu et de St Antoine, je cite ainsi dans le désordre Aurélie, Juliette, Fatia, Philippe, JLP, Beghdad, Jef, Jennifer, Carl, Emmanuel, PA, PC, Muriel, Erwan, Dominique, Christelle, Céline, Diane, Barbara, Félicité, Jessica, Oriane, Nicolas, Laetitia, Monik et j'en oublie encore probablement énormément.

Un très très grand merci à ma famille, mes parents, mon frère et maman bis (qui a force fait partie de la famille) pour leur soutien de tous les instants, dans cette thèse de doctorat, mais aussi au quotidien, dans les moments joyeux ou ceux plus difficiles.

Enfin, merci à Anne-Laure, Sofia et Sacha. Merci de votre soutien, votre aide et votre amour.

Résumé

Les essais contrôlés randomisés et les revues systématiques qui les synthétisent sont des éléments indispensables à la pratique de la médecine fondée sur les preuves. Celle-ci accorde un poids très important à la recherche biomédicale, et sera donc compromise si la base des éléments de preuve s'avère erronée. La mauvaise qualité de la recherche biomédicale est dénoncée depuis plusieurs années, mais en 2009, I.Chalmers et P.Glasziou intègrent ces critiques dans un concept plus global, le gaspillage de la recherche. Ils affirment que sans des rapports de recherche accessibles, honnêtes et utilisables, la recherche biomédicale ne pourra pas aider les patients et soignants à prendre des décisions éclairées, et pourrait ainsi être considérée comme gaspillée. Une recherche mal conçue, mal réalisée, non ou mal rapportée serait ainsi gaspillée. Ils ont estimé que près de 85 % de la recherche biomédicale le serait.

Dans le cadre de cette thèse, nous nous sommes intéressés au gaspillage, évitable, de la recherche dans les essais cliniques. Nous nous sommes plus spécifiquement concentrés sur le gaspillage lié à une mauvaise planification des essais (choix des méthodes et des critères de jugement) ou à la présentation sélective ou incomplète des critères de jugement. Puis nous avons estimé dans quelle mesure ce gaspillage pouvait être évité de manière simple et peu coûteuse.

Nos travaux suggèrent que 1) que des ajustements méthodologiques simples et peu coûteux permettraient de limiter le risque de biais dans 50 % des essais, et ainsi réduire partiellement le gaspillage de la recherche, et 2) que de nombreux essais ne mesuraient ou ne rapportaient pas complètement les critères de jugement importants, mais que ce gaspillage aurait pu être partiellement évité pour la majorité des essais.

Mots-clés : Biais ; critères de jugement ; essai contrôlé randomisé ; gaspillage de la recherche ; qualité, épidémiologie ; revues systématiques ; méta-analyses.

Abstract

Title: Avoidable waste of research in clinical trials.

Randomized controlled trials and systematic reviews are essential for the practice of evidence-based medicine (EBM). EBM will be compromised if the evidence issued from biomedical research proves to be biased. Many authors previously denounced the poor quality of research, but in 2009 I.Chalmers and P.Glasziou integrate these criticisms into a more global concept, waste of research. It could be defined as research that fails to help patients and their clinicians to make informed decisions as without accessible, honest and usable reports, research cannot help research end users. They estimated that it could represent up to 85% of health research.

In this thesis, we focused on the avoidable waste of research in clinical trials. We had a particular interest in waste of research due to poor trial planning (trials methods or choice of outcomes) or to selective and incomplete reporting of outcomes. Then we estimated to what extent this waste could have been avoided by simple and inexpensive methodological adjustments.

Our results suggest that 1) simple and inexpensive methodological adjustments could have limited the risk of bias in 50% of clinical trials, and thus partially reduce the burden of waste of research, and 2) that many trials did not measure or report completely the important outcomes, but that this waste could have been partially avoided for the majority of the trials.

Keywords: Bias; outcomes; randomised controlled trial; waste of research; quality, epidemiology; systematic reviews; meta-analyses.

Table des matières

Laboratoire d'accueil	9
Publications liées à la thèse.....	10
Liste des principales abréviations	11
Listes de figures et tableaux.....	12
Listes des tableaux.....	13
1. Introduction	15
1. 1. Médecine fondée sur les preuves	15
1.1.1 Principes de la médecine fondée sur les preuves.....	15
1.1.2 Hiérarchie des preuves	16
1.1.3 Revues systématiques et méta-analyses	18
1.1.4 Limites des revues systématiques et méta-analyses.....	20
1. 2. Le gaspillage de la recherche	22
1.2.1 Choix des questions de recherche non pertinentes.....	24
Besoins des utilisateurs de la recherche	25
Absence de prise en compte des éléments de preuve existants.....	27
1.2.2 Mauvaise planification des essais cliniques : choix des méthodes et critères de jugement.....	30
Méthodes inadéquates dans les essais cliniques	30
Choix des critères de jugement	37
1.2.3 Accessibilité des résultats de la recherche	40
Non-publication des résultats d'un essai clinique.....	40
1.2.4 Rapports de recherche transparents, non biaisés et complets.....	45
Présentation sélective ou incomplète des résultats	45
Exhaustivité des rapports de recherche.....	48
Embellissement des résultats	51
1.3 Justification de la démarche de thèse	53
2. Gaspillage évitable de la recherche dû à l'emploi de méthodes inadéquates dans les essais cliniques	55
2.1. Justification de l'étude.....	55
2.2. Objectifs.....	55
2.3. Méthodes.....	55
2.4. Résultats.....	58

2.5. Discussion.....	62
2.6. Article.....	63
3. Gaspillage, évitable, de la recherche lié aux critères de jugement dans les essais cliniques	81
3.1 Justification de l'étude	81
3.2. Objectifs	82
3. 3. Méthodes	82
3. 4. Résultats.....	88
3. 5. Discussion.....	95
3.1.6. Articles	97
4. Discussion et perspectives	125
4. 1. Discussion.....	125
4.1.1. Principaux résultats.....	125
4.1.2. Implications	127
4.1.3. Limites.....	135
Impact de la méthode d'évaluation du risque de biais	135
Impact de la source des données : revues systématiques Cochrane	136
Impact de l'évaluation par les experts	137
Impact du choix des critères de jugement dit importants	138
4. 2. Conclusion	139
4. 3. Perspectives.....	140
Gaspillage de la recherche dans les essais cliniques : priorisation et critères de jugement	140
Gaspillage de la recherche dans les revues systématiques	141
Références	144

Laboratoire d'accueil

Centre de Recherche Épidémiologie et Statistiques Sorbonne Paris Cité (CRESS) – U1153,
équipe METHODS : Méthodes en Évaluation Thérapeutique des Maladies Chroniques

Centre d'Épidémiologie Clinique

Hôpital Hôtel Dieu 1 place du parvis Notre Dame 75004 Paris, France

Publications liées à la thèse

Articles publiés :

1. **YORDANOV, Y** ; Dechartres, A ; Porcher, R ; Boutron, I ; Altman, DG ; Ravaud, P, *Avoidable waste of research related to inadequate methods in clinical trials.*, BMJ, 2015, 350, h809
2. **YORDANOV, Y** ; Dechartres, A ; Ravaud, P, *Patient-important outcomes in systematic reviews: Poor quality of evidence.*, PLoS One, 2018, 13, e0195460
3. **YORDANOV, Y** ; Dechartres, A ; Atal, I ; Tran, VT ; Boutron, I ; Crequit, P & al, *Avoidable waste of research related to outcome planning and reporting in clinical trials.*, BMC Med, 2018, 16, 87

Liste des principales abréviations

Années de vie ajustées à l'incapacité : DALY

British Medical Journal: BMJ

Critères de jugement importants pour les patients : CJIP

Critères de jugement importants : CJI

Comparative effectiveness research: CER

Consolidated Standards of Reporting Trials: CONSORT

Consort-based WEB tool: COBWEB

Core Outcome Set: COS

Evidence Based Medicine : EBM

Enhancing the QUALity and Transparency Of health Research: EQUATOR

Essais contrôlés randomisés : ECR

Food and Drug Administration: FDA

Hydroxyéthylamidon: HEA

International Consortium for Health Outcome Measurement: ICHOM

James Lind alliance : JLA

Patient-Centered Outcomes Research Institute: PCORI

Priority Setting Partnership: PSP

Risk of bias: RoB

Summary of Findings: SoF

Listes de figures et tableaux

Figure 1. Pyramide des niveaux de preuves.

Figure 2. Étapes et principales causes du gaspillage de la recherche

Figure 3. Proportion d'essais contrôlés randomisés dans un domaine donné et des besoins de santé correspondant, en années de vie ajustées à l'incapacité (DALY).

Figure 4. Méta-analyses conventionnelles et cumulatives de 33 essais contrôlés randomisés évaluant la streptokinase intraveineuse dans la prise en charge de l'infarctus aigu du myocarde.

Figure 5. Exemple de graphique rapportant les résultats de l'évaluation du risque de biais dans une revue Cochrane.

Figure 6. Taux de publication d'un échantillon aléatoire de 677 essais cliniques enregistrés sur ClinicalTrials.gov de 2000 à 2007.

Figure 7. Diagramme en forêt illustrant le problème des essais non publiés dans le domaine de la prise en charge de la douleur chronique.

Figure 8. Géométrie du réseau pour les essais rapportant un effet du traitement pour A) douleur, B) nausée, C) vomissement et D) effets indésirables dans une méta-analyse en réseau portant sur les analgésiques non opiacés dans la douleur post-opératoire.

Figure 9. Pourcentage d'études présentant une description suffisante de l'intervention initialement (sur la seule base du document publié) et après l'obtention d'informations supplémentaires auprès des auteurs.

Figure 10. Sélection des revues et essais cliniques, et évaluation du risque de biais.

Figure 11. Gaspillage évitable de la recherche liée aux méthodes inadéquates dans les essais avec au moins un domaine confirmé à risque élevé de biais (n = 142).

Figure 12. Organigramme du processus de sélection.

Figure 13. Critères de jugement important pour les patients dans des revues Cochrane récentes

Figure 14. Gaspillage évitable de la recherche lié aux critères de jugement importants, manquants, dans les ECRs publiés en 2010 ou plus tard (n = 291).

Listes des tableaux

Tableau 1. « Risk Of Bias tool » (RoB) de la Collaboration Cochrane

Tableau 2. Présentation tabulaire de l'évaluation du risque de biais d'un essai contrôlé randomisé

Tableau 3. Raisons de l'absence des critères de jugement importants des essais cliniques

1. Introduction

1. 1. Médecine fondée sur les preuves

1.1.1 Principes de la médecine fondée sur les preuves

Depuis l'introduction de la médecine expérimentale et scientifique, le médecin se différencie du guérisseur par son aptitude à douter et par sa capacité à ne pas considérer la guérison pour preuve (1-3). Plus de 25 ans après sa première description, il est difficile d'imaginer un système de soins moderne qui ignorerait l'Evidence Based Medicine (EBM) ou médecine fondée sur les preuves. Ce qui a d'abord été introduit en 1992 comme une nouvelle méthode pédagogique a provoqué depuis un changement profond dans tous les champs de la santé (4). Elle a été définie, et décrite, par David Sackett comme une médecine qui utiliserait de manière explicite, consciencieuse et judicieuse les meilleures preuves disponibles pour la prise en charge individualisée de chaque patient (4-6).

La pratique de l'EBM implique de conjuguer 3 éléments (6, 7) :

1. L'expertise clinique individuelle : les connaissances acquises par expérience
2. Les préférences, valeurs et désirs du patient
3. Les meilleurs éléments de preuves disponibles : les données de la recherche scientifique.

L'avènement de l'EBM a permis de formaliser la nécessité de fonder ses décisions sur les meilleurs éléments de preuves disponibles. Celles recherchées dans ce cadre doivent provenir des études au meilleur niveau de preuve scientifique, dans les domaines du diagnostic, de la thérapeutique, de la prévention, ou encore du pronostic. Or, dans certaines situations, parmi les éléments de preuves on retrouve des extrapolations approximatives de principes physiopathologiques (8), ou des croyances basées sur des mécanismes d'actions non plausibles ni scientifiquement démontrables (9). Ainsi toutes les preuves ne sont pas égales, il convient de les hiérarchiser (10).

1.1.2 Hiérarchie des preuves

Le niveau de confiance accordé aux éléments de preuves disponibles va dépendre du schéma de l'étude dont elles proviennent et de la manière dont celle-ci a été menée. L'essai contrôlé randomisé (ECR) a régulièrement été décrit comme le schéma d'étude le plus à même de limiter les biais, et d'apporter un haut niveau de preuves (11, 12). Ainsi des résultats issus d'un ECR seront considérés comme plus solides, et plus fiables, que ceux provenant de séries de cas ou d'études observationnelles. En médecine, les ECRs ont longtemps été interprétés de manière isolée bien que la première évocation des techniques permettant de combiner les résultats de plusieurs études entre elles remonte à plus d'un siècle (13). Dans ce rapport discutant de travaux portant sur la fièvre typhoïde, K.Pearson souligne les limites de l'interprétation isolée des études : « Many of the groups (...) are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved » (13). En 1979, un

chercheur anglais, Archibald Cochrane, appelle la profession médicale à préparer des résumés critiques, régulièrement mis à jour, de l'ensemble des essais cliniques pertinents publiés dans chaque spécialité ou sous-spécialité médicale : « It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials » (14). Dès l'instant qu'il existe des données, il conviendrait de les rassembler avec rigueur au sein de revues systématiques, celles-ci pouvant inclure une synthèse quantitative des résultats des différents essais : la méta-analyse (14-17). Les premières revues systématiques et méta-analyses ont été publiées dans les années 1980 et se sont considérablement multipliées par la suite, notamment grâce à la Collaboration Cochrane, créée par Sir Ian Chalmers en réponse à l'appel d'Archibald Cochrane(14).

A été ainsi établie une pyramide des niveaux de preuves, ayant pour base les séries de cas cliniques ou encore les études cas-témoins et au sommet de laquelle on retrouve les revues systématiques d'essais contrôlés randomisés, avec ou sans méta-analyses (fig. 1) (18, 19).

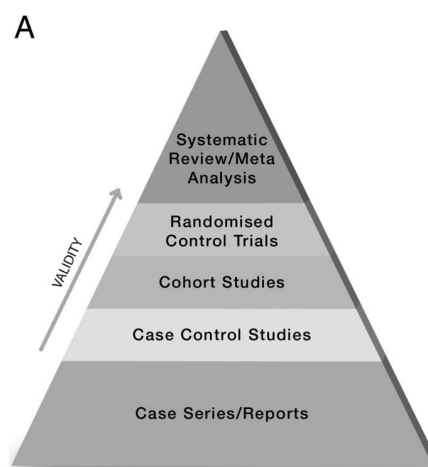


Figure 1. Pyramide des niveaux de preuves. Figure extraite de Murad *et al*(19)

1.1.3 Revues systématiques et méta-analyses

Le nombre d'essais cliniques publiés annuellement est passé d'environ 10 000 en 1990, à près de 25 000 en 2007 et plus de 29 000 en 2017. Ceci correspond donc, selon l'évaluation de H.Bastian *et al*, à près de 75 nouveaux essais contrôlés randomisés publiés chaque jour (20). Le volume de connaissances disponibles croissant de manière exponentielle, il est devenu impossible pour les professionnels de santé de lire, analyser et synthétiser individuellement les données disponibles afin de maintenir à jour leurs connaissances médicales (20, 21). Les essais cliniques sont donc régulièrement synthétisés dans des revues systématiques et des méta-analyses. Ces revues permettent d'examiner de manière systématisée l'ensemble des données disponibles, d'en analyser l'hétérogénéité, la robustesse, les biais éventuels et ainsi de conclure à l'efficacité, ou non, d'une thérapeutique.

Elles sont aussi essentielles aux chercheurs, tout essai clinique devant impérativement commencer par une revue systématique de l'ensemble de la littérature disponible afin d'en justifier l'intérêt (22). Hormis motiver la nécessité, ou futilité, de mener l'essai, la revue systématique préalable doit permettre aux auteurs de replacer leurs résultats dans le contexte de l'ensemble des données existantes, tel que recommandé par la déclaration CONSORT (Consolidated Standards of Reporting Trials) (23-25). Certains journaux, tels que le *Lancet* et les journaux du groupe *Lancet*, exigent depuis 2010 que tout rapport d'essai contrôlé randomisé mène et rapporte, dans un encadré spécifique de la section discussion, une revue systématique de la littérature de la question de recherche abordée (26). De manière plus controversée, certains auteurs suggèrent même qu'un nouvel essai ne devrait être entrepris qu'à la condition que la méta-analyse des données existantes soit peu concluante et que les résultats éventuels de l'essai en cours de planification puissent modifier la méta-analyse, une fois celle-ci mise à jour (27, 28).

Une revue systématique, contrairement à une revue narrative, est une revue qui va utiliser une méthode systématique afin de minimiser les biais et ainsi d'augmenter la fiabilité et la précision des résultats (29, 30). Les revues systématiques reposent sur des méthodes explicites, rigoureuses et reproductibles (31, 32) afin de répondre à une question clinique précise.

Une revue systématique répond aux caractéristiques essentielles suivantes :

- 1) Une question clinique claire, qui va donc impliquer une définition précise des objectifs ;
- 2) Des critères d'éligibilité pré défini, à partir des objectifs ;
- 3) Une recherche systématique visant à identifier l'ensemble des études répondant aux critères d'éligibilité, qu'elles aient été publiées ou non ;
- 4) Une évaluation de la validité interne des résultats des études ;
- 5) Une présentation systématique de l'ensemble des caractéristiques et des résultats des études sélectionnées.

Les auteurs de revues systématiques peuvent éventuellement synthétiser de manière quantitative leurs résultats à l'aide de méta-analyses. Il s'agit d'un outil statistique qui permet de combiner les résultats de plusieurs études en une estimation unique de l'effet traitement (33, 34). La méta-analyse est une moyenne de l'effet traitement estimé dans les études individuelles, pondérée selon la précision des effets des thérapeutiques testés dans les essais individuels. Les études avec l'effet le plus précis auront ainsi un poids plus important dans la méta-analyse et contribueront davantage à l'estimation combinée. La majorité des méta-analyses sont réalisées à partir des données agrégées issues de la littérature, c'est-à-dire des résultats tels que rapportés dans les rapports d'essais cliniques. Une autre possibilité consiste à utiliser les données

individuelles des patients, obtenues auprès des auteurs de l'essai clinique, et ainsi réaliser une méta analyse sur données individuelles. Ce type de méta-analyses est considéré comme le gold standard des méta-analyses et devient de plus en plus courant (35).

1.1.4 Limites des revues systématiques et méta-analyses

L'expression « garbage in, garbage out » décrit parfaitement le besoin d'alimenter les revues systématiques d'essais cliniques de qualité. Un exemple caricatural a été publié par Zarychanski *et al* (36). Ils rapportent une revue systématique de l'utilisation des hydroxyéthylamidons (HEA) comme soluté de remplissage vasculaire pour des patients en réanimation ou en soins intensifs. Les HEA ont longtemps été suspectés d'avoir des effets secondaires graves. Dans cet article, les auteurs rapportent une méta-analyse incluant 38 ECR qui ne permet pas de conclure sur le risque de décès de toutes causes liées à l'utilisation des HEA. Mais, parmi les 38 essais, 7 ont été conduits par le Dr Joachim Boldt, anesthésiste allemand dont tous les essais réalisés à partir de 1999 ont été rétractés suite à une enquête ayant mis en évidence des fabrications de données. Or, retirer ces essais, sur lesquels pèse un fort soupçon de fraude, modifie complètement la conclusion de la méta-analyse avec un risque relatif de mort qui devient alors statistiquement significatif (RR=1,09, IC 95 % 1,02 - 1,17), malgré le poids faible de ces 7 essais dans la méta-analyse de la mortalité liée au HEA. Un autre exemple, tout aussi caricatural, a été décrit dans l'évaluation des thérapeutiques préventives des nausées post-opératoires par le chercheur Japonais Yoshitaka Fujii (183 articles rétractés) (37).

Les revues systématiques, et les méta-analyses sont par nature rétrospectives et dépendent des données issues des études individuelles incluses. Les fraudes, falsifications ou manipulations des données sont des pratiques rares et extrêmes, mais de manière plus commune de nombreux biais sont susceptibles de remettre en cause la validité des essais. Or, si les revues incluent des études biaisées leurs résultats risquent d'être biaisés, et ainsi de conduire à des conclusions incorrectes.

Il a été rapporté que les études avec des résultats statistiquement significatifs étaient publiées plus rapidement (38, 39), plus fréquemment (40-50), et souvent plusieurs fois (51-53) ou encore étaient plus fréquemment citées (54-56) que les études ne rapportant pas de différence significative. A été également mise en évidence une présentation sélective des résultats des ECRs, qui consiste à sélectionner les résultats qui seront rapportés selon leur significativité statistique (57-59). Donc, le risque de retrouver et d'inclure plus d'études et de résultats statistiquement significatifs est bien réel. Ceci est susceptible de biaiser les résultats des méta-analyses. C'est pourquoi l'évaluation de la validité des ECRs est l'une des étapes essentielles des revues systématiques et méta-analyses.

1. 2. Le gaspillage de la recherche

L'EBM accorde, par essence, un poids très important à la recherche biomédicale et à la littérature scientifique. Mais elle devient inutile, voire dangereuse, si la base des éléments de preuve s'avère erronée ou compromise. Ainsi, affirmer la primauté de la médecine basée sur les preuves d'un côté et avoir une recherche biomédicale de qualité médiocre de l'autre, mettrait les chercheurs et les cliniciens face à un important dilemme (60-62). À lire les éditeurs du *Lancet* (R.Horton) ou du *New England Journal of Medicine* (M.Angell) la recherche scientifique traverse une crise majeure. Il ne serait plus possible de croire à une grande partie de la recherche clinique publiée, une grande proportion, peut-être la moitié, serait tout simplement fausse (63, 64). Ce constat préoccupant n'est pour autant pas récent. Il y a plus de 20 ans, et de manière quasi concomitante aux premières descriptions de l'EBM(4-6), Douglas Altman publiait dans le *British Medical Journal* un éditorial réquisitoire sur le scandale de la mauvaise recherche biomédicale (65).

« Que penser d'un médecin qui utiliserait le mauvais traitement, volontairement ou par ignorance, ou qui utiliserait mal le bon traitement ? Un tel comportement serait considéré comme contraire à l'éthique, non professionnel et inacceptable.

Que penser des chercheurs qui utilisent les mauvaises techniques, utilisent à tort les bonnes, interprètent et rapportent mal les résultats, citent la littérature de manière sélective et tirent des conclusions injustifiées ? Nous devrions être consternés.

Nous avons besoin de moins de recherche, d'une recherche de meilleure qualité et d'une recherche effectuée pour les bonnes raisons » (65).

— Douglas Altman (1948-2018) —

Cette même littérature scientifique qui sert de fondations à l'EBM, regorgerait d'études mal ou partiellement rapportées, avec de faibles effectifs, des statistiques approximatives, de petits effets traitement ou encore des conflits d'intérêts flagrants, et serait pour moitié tout simplement fausse (63, 65-67). La mauvaise qualité de la recherche biomédicale est dénoncée depuis plusieurs années, mais en 2009, I.Chalmers et P.Glasziou intègrent ces critiques dans un concept plus global, le gaspillage de la recherche (22). En partant de l'exemple personnel d'un chercheur qui s'est retrouvé dans la position de patient, ils affirment que toute recherche qui échoue à aider les patients et leurs médecins à prendre des décisions éclairées pourrait être considérée comme gaspillée. Une recherche mal conçue, mal réalisée, non ou mal rapportée serait ainsi gaspillée. Ils ont estimé que jusqu'à 85 % de la recherche biomédicale pourrait être considérée comme gâchée (22).

Dans une récente série du *Lancet*, il a été décrit que ce gaspillage affecterait tous les stades de la recherche (fig. 2), allant du choix de questions de recherche non pertinentes, à la non présentation ou présentation sélective des résultats des essais cliniques (68-79).

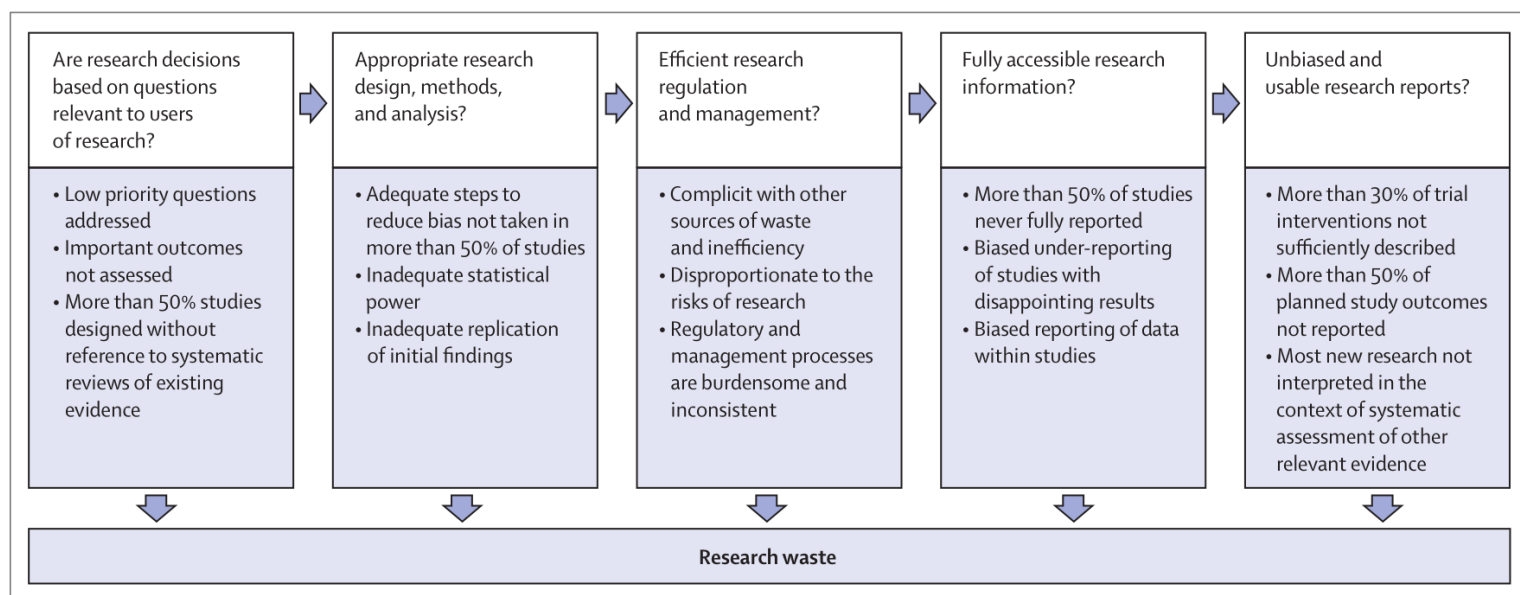


Figure 2. Étapes et principales causes du gaspillage de la recherche (22, 79). Figure extraite de Macleod *et al*(79)

1.2.1 Choix des questions de recherche non pertinentes

Un système de recherche efficace doit être capable de répondre aux problèmes de santé importants pour les patients et pour la société. Il doit se baser sur les besoins de l'ensemble des utilisateurs de la recherche (patients, professionnels de santé, décideurs des politiques de santé publique et financeurs) pour définir des priorités de recherche en tenant compte des éléments de preuve existants.

Besoins des utilisateurs de la recherche

Pour les chercheurs, il existe un impératif moral à prendre en compte les besoins et priorités de ceux qui pourraient profiter des bénéfices de la recherche (80). Tenir compte de la perspective des patients dans la recherche en élargirait la portée et aiderait à répondre aux questions susceptibles de provoquer les plus grandes améliorations en termes de santé publique. Pourtant, un large fossé semble exister entre la recherche financée et conduite, et ce que les patients, les soignants et le grand public aimeraient voir évaluer (81-83). La « James Lind Alliance » (JLA) est une institution qui réunit des patients, des soignants et des chercheurs pour identifier et hiérarchiser les incertitudes ainsi que les « questions sans réponses » dans le champ de l'évaluation thérapeutique, au cours de « Priority Setting Partnership » (PSP) (84). Une revue menée par la JLA ayant analysé 374 études traitant de la priorisation de la recherche a rapporté que seulement 9 études (2,5 %) comparaient les vues des patients et/ou des soignants à celles des chercheurs (85). Dans une analyse de la littérature publiée et non publiée sur l'arthrose de genou, les éléments de preuves disponibles étaient par exemple très largement dominés par des études d'interventions pharmacologiques (n=550, 59 %) par rapport aux essais évaluant des interventions non pharmacologiques (chirurgie) (238, 26 %) (83). Mais pour plus de la moitié des usagers de la recherche interrogés (rhumatologues, médecins généralistes, kinésithérapeutes et patients), celle-ci devait en priorité s'intéresser à l'éducation thérapeutique et aux thérapeutiques non médicamenteuses (83). Près de 20 ans plus tard, cette inadéquation reste d'actualité (82). Dans le domaine de la recherche en oncologie, les deux domaines prioritaires selon les patients (soins palliatifs, prévention et diagnostic précoce) représentaient moins de 15 % des investissements financiers de l'« UK Cancer Research Institute » (86). Les patients estiment que la recherche sur les thérapeutiques pharmacologiques est une priorité dans 18 % des « Priority Setting Partnership », mais celle-ci représenterait 86 % (689/798) des essais

industriels et 37 % (397/1069) des essais cliniques menés avec une autre source de financement (82).

Hormis le choix de la question de recherche qui est en inadéquation avec les besoins des patients et des soignants, il a été régulièrement décrit que le financement de la recherche correspondait peu au fardeau réel de la maladie et/ou aux besoins des populations (87-95). Dans une étude épidémiologique portant sur un ensemble d'essais cliniques, Emdin *et al* retrouvaient une faible corrélation entre le fardeau des maladies selon la classification de l'Organisation mondiale de la Santé et le nombre d'essais cliniques publiés ou le nombre de patients randomisés (92). D'autres chercheurs ont évalué l'adéquation entre la recherche menée et les besoins en termes de santé des populations en prenant l'exemple de 2 priorités de santé publique : la consommation de tabac et l'infection à VIH (95). En ce qui concerne le tabagisme, 4 % des essais inclus dans les revues systématiques et 2 % des essais en cours ont été réalisés dans des pays qui représentaient 70 % de la mortalité liée au tabagisme. Pour l'infection à VIH, respectivement 31 % et 33 % des essais publiés ou en cours ont été réalisés dans des pays qui représentaient 99 % de la mortalité (95). De manière comparable, les pathologies fortement prévalentes dans les pays développés auraient ainsi fait l'objet de sept à huit fois plus d'essais cliniques que celles affectant majoritairement les pays en voie de développement (91). L'analyse de 117 180 essais cliniques, ayant inclus 42,6 millions de patients, retrouvait un déséquilibre entre le besoin de santé des populations (exprimés en années de vie ajustées à l'incapacité) et le volume de recherche disponible (fig. 3) (90). Ce déséquilibre affectait en particulier les pays en voie de développement et la recherche sur les maladies néonatales ou encore sur les pathologies infectieuses communes (dengue, bilharziose).

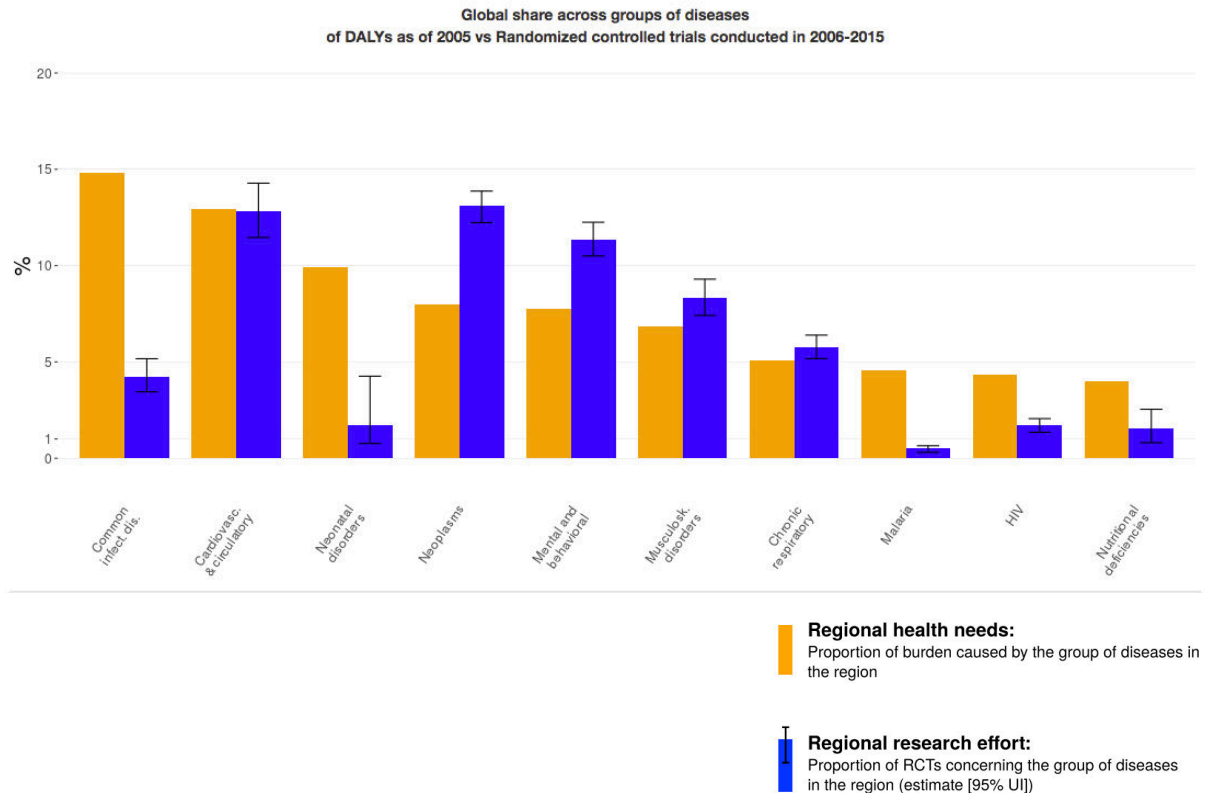


Figure 3. Proportion d’essais contrôlés randomisés dans un domaine donné et des besoins de santé correspondant, en années de vie ajustées à l’incapacité (DALY). Figure extraite d’Atal *et al*(90)

Absence de prise en compte des éléments de preuve existants

Afin d’aider les chercheurs à identifier et prioriser les besoins à évaluer, un examen approfondi des éléments de preuves existants est indispensable (96-98). Une telle évaluation permet de dresser un panorama de la recherche déjà menée, et devrait ainsi permettre une meilleure adéquation de la recherche à mener avec celle préexistante. Mais ce préalable essentiel semble insuffisamment pris en compte (99-103).

Il est toujours possible de discuter le point où la réplication de travaux antérieurs devient une redondance futile et inutile, mais pour éclairer cette discussion il est nécessaire d'avoir recours à une évaluation systématique et méthodique de la littérature existante (104, 105). Les revues systématiques, avec ou sans méta-analyses, sont des outils d'une grande importance dans ces situations, car ils permettent de mettre en évidence des effets thérapeutiques, ou néfastes que des essais individuels peinaient à exposer (99, 106-109). Mais cet exercice n'est que trop rarement mené avant de démarrer les essais cliniques, malgré les recommandations d'experts et de groupes de journaux (23-26). Une analyse de 50 études, ayant inclus plus de 1500 méta-analyses cumulatives, met en évidence que dans de nombreuses situations les effets bénéfiques (ou néfastes) de traitements auraient pu être identifiés plus tôt qu'ils ne l'ont été si les chercheurs avaient systématiquement recherché les éléments existants (110). Par exemple, depuis 2002, l'effet de l'acide tranexamique sur le risque de saignement en chirurgie cardiaque semblait connu et stable. Pourtant, entre 2003 et 2011, 24 essais cliniques, ayant inclus plusieurs centaines de patients, ont été menés (99). De manière similaire, la méta-analyse cumulative, rapportée par Fergusson *et al*, sur l'aprotinine en chirurgie cardiaque retrouvait un effet traitement stable et très en faveur de l'intervention à la douzième étude, datant de juin 1992 (Odds Ratio= 0,25, P <0,000 001) (100). À cette date 2500 patients avaient déjà été randomisés. Les essais cliniques ont continué, le dernier essai clinique inclus dans cette méta-analyse date de juin 2002, et plus de 5500 patientes supplémentaires avaient été randomisées. Des résultats comparables ont été retrouvés pour l'évaluation de l'effet des bêtabloquants ou encore de la thrombolyse sur la mortalité des patients souffrant d'un infarctus du myocarde (fig. 4) (106, 109).

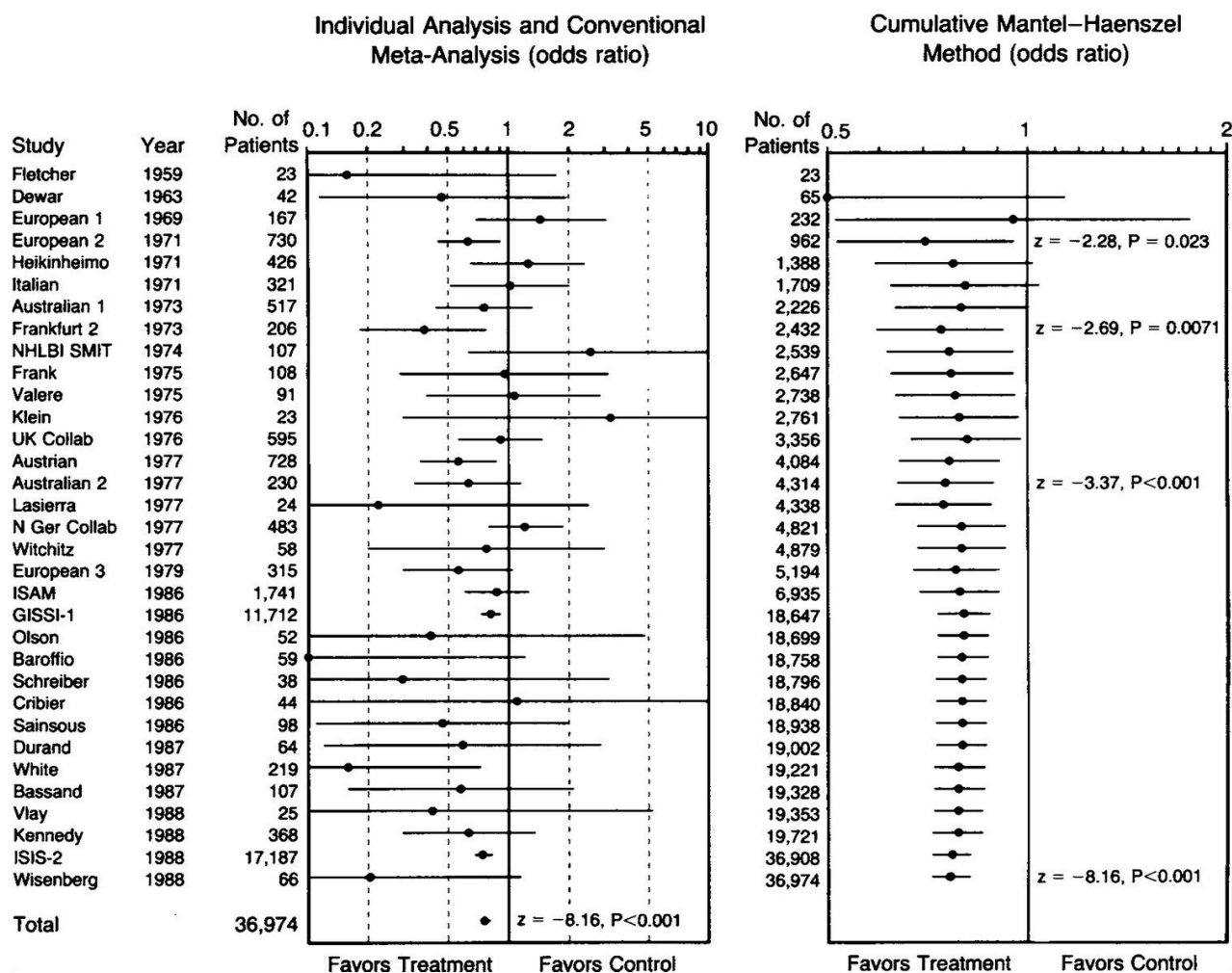


Figure 4. Méta-analyses conventionnelles et cumulatives de 33 essais contrôlés randomisés évaluant la streptokinase intraveineuse dans la prise en charge de l'infarctus aigu du myocarde. Figure extraite de Lau *et al*(109)

Une recherche systématique de l'ensemble des preuves disponibles, dans ces situations, aurait permis non seulement de réduire le gaspillage de la recherche secondaire à la conduite d'essais cliniques futiles, mais aussi vraisemblablement de réduire la morbidité et parfois la mortalité, à la fois chez les patients affectés à des traitements relativement moins efficaces ou réellement nocifs dans des essais inutiles, et chez les patients en général.

Moins de la moitié des auteurs d'essais contrôlés randomisés (ECR) rapportent s'être aidés de revues systématiques pour la planification de leurs études (103), et près de la moitié semblaient ignorer l'existence de revues systématiques dans le champ de leur essai clinique (102). Ces résultats semblent confirmés par l'analyse des protocoles d'ECR qui par exemple, dans moins d'1 % des cas, se sont appuyés sur une méta-analyse des résultats préexistants pour estimer la taille de leur échantillon (111). Cette absence d'analyse systématique de la littérature entraîne des citations partielles et biaisées de la recherche (54, 112). Dans une analyse ayant porté sur près de 40 années de publications scientifiques, Robinson *et al* rapportent que moins d'un quart des études précédant un essai clinique étaient citées dans les rapports de celui-ci, avec une médiane de 2 essais cités (101). Les études antérieures ayant rapporté des résultats positifs étant plus souvent citées que celles rapportant des résultats négatifs (54). Ces citations partielles de la littérature participeraient à une mauvaise estimation de la taille de l'échantillon et donc à des études sous ou surdimensionnées. Une des conséquences de ce biais d'optimisme est, par exemple dans le domaine du cancer, que seulement 17 % des essais avaient un effet traitement correspondant aux attentes originales des chercheurs (113).

1.2.2 Mauvaise planification des essais cliniques : choix des méthodes et critères de jugement

Méthodes inadéquates dans les essais cliniques

Pour qu'un essai clinique puisse être utile aux utilisateurs de la recherche, il doit présenter des résultats valides (67, 77). On distingue habituellement la validité interne de la validité externe.

La validité interne d'un essai est dépendante des choix méthodologiques des auteurs et de la capacité de ceux-ci à limiter les risques de biais. La validité externe dépend de la capacité des cliniciens à transposer les résultats de l'essai dans leur pratique quotidienne (114-117). La Collaboration Cochrane recommande de se concentrer sur la validité interne des essais plutôt que la validité externe (29).

Depuis de nombreuses années, nous savons, grâce aux études métaépidémiologiques, que certaines caractéristiques des essais cliniques risquent de biaiser les résultats et avoir ainsi une influence sur l'ampleur des estimations des effets traitements des interventions évaluées (118-135). La première étude métaépidémiologique a été publiée en 1995 et analysait 33 méta-analyses, ayant inclus 250 essais (125). Elle a évalué l'influence de l'assignation secrète, de l'insu (étude en aveugle ou ouverte) et de l'exclusion de patients de l'analyse sur l'effet traitement. L'assignation secrète du traitement correspondant aux moyens mis en œuvre pour que l'assignation des groupes de traitement soit imprévisible (par exemple par l'emploi d'enveloppes scellées, numérotées et opaques ou d'une randomisation centralisée). Les résultats retrouvaient des effets traitements plus importants en moyenne dans les études avec une assignation secrète inadéquate, ou non rapportée, par rapport aux études avec une assignation secrète adéquate. Des résultats similaires étaient observés pour les essais décrits comme en insu par rapport à ceux dits ouverts. Depuis, de nombreuses autres études ont retrouvé des résultats montrant des estimations exagérées dans les essais cliniques avec une génération inadéquate de la séquence de randomisation (127-129), une assignation secrète inadaptée, une absence d'insu (en particulier dans les essais avec des critères de jugement subjectifs) (125, 127, 132), ou encore ayant exclus des patients de l'analyse (130, 131).

En 2005, Hewitt *et al* y rapportaient que dans plus de 40 % des essais publiés dans les 4 principales revues générales de médecine, l'assignation secrète était soit inadéquate soit mal rapportée (133). L'exclusion de patients des analyses était également très fréquente, avec près de 54 % des essais évalués qui excluaient au moins une partie des patients (136). Ce qui laisse à penser qu'un grand nombre d'essais seraient susceptibles d'être biaisés. Ces biais liés aux caractéristiques intrinsèques des études se retrouvent également dans d'autres domaines que l'évaluation thérapeutique. Par exemple, dans une étude métaépidémiologique de 487 études évaluant des tests diagnostiques, l'utilisation de références différentes pour la positivité des tests, mais encore l'absence de lecture en double insu des résultats surestimait les performances diagnostiques de ceux-ci (137).

Cela impose de rechercher et d'évaluer la présence de possibles biais lors de l'analyse de la validité d'un essai. La Collaboration Cochrane recommande d'évaluer au moins 4 grands types de biais lors de l'analyse d'un ECR (29) :

- Le biais de sélection : défini par la présence de différences systématiques entre les individus des groupes à comparer. Le risque de biais de sélection s'évalue en analysant la randomisation ainsi que l'assignation secrète. Pour limiter le risque de biais de sélection, il est nécessaire de générer une séquence aléatoire imprévisible puis de la mettre en œuvre de manière à masquer les traitements jusqu'à ce que les patients aient été officiellement affectés à leurs groupes.
- Le biais de performance : défini par la présence de différences systématiques dans l'attention et les soins accordés entre les participants. Le risque de biais de performance s'évalue en analysant l'insu, en particulier des participants et du personnel soignant.

- Le biais de détection : défini par la présence de différences systématiques dans l'évaluation du critère de jugement. Ce biais est particulièrement important quand il s'agit de critères de jugement subjectifs, notamment de critères de jugement rapportés par les patients (129). Le risque de biais de détection s'évalue en analysant l'insu des évaluateurs des critères de jugement. Il est important de l'évaluer en fonction du type de critère de jugement.
- Le biais d'attrition : défini par la présence de différences systématiques quant aux perdus de vue. Le risque de biais d'attrition s'évalue en analysant le nombre et les raisons des perdus de vue, et peut-être limités par les auteurs des ECRs par l'utilisation d'analyses en intention de traiter (ITT) avec une gestion appropriée des données manquantes.

La Collaboration Cochrane, en se basant sur les résultats des études méta épidémiologiques, a développé le «Risk Of Bias tool» (RoB) et émis des recommandations sur les modalités d'évaluation de la validité interne des essais cliniques à l'aide de ce dernier (138). Il comporte 4 domaines principaux, liés théoriquement et empiriquement au risque de biais dans les ECRs. Les domaines sont nommés selon les différents biais correspondant : biais de sélection (méthode de génération de la séquence de randomisation et assignation secrète), biais de performance (aveugle des participants et du personnel), biais de détection (aveugle de l'évaluateur) et le biais d'attrition (données manquantes). Il existe également deux autres biais que les chercheurs sont invités à évaluer : le biais de présentation sélective des critères de jugement et les autres risques de biais. L'évaluation de chaque domaine se fait en 2 étapes : description de ce qui est rapporté dans l'article et jugement du risque de biais, en classant celui-ci comme faible, incertain ou enfin élevé (tableau.1) (138).

Domaine	Description	Jugement
Génération de la séquence de randomisation	Décrire la méthode de génération de la séquence de randomisation, avec assez de détails pour évaluer si elle permet d'obtenir des groupes comparables	Biais de sélection (allocation des interventions biaisée) en raison de la génération d'une séquence de randomisation inadéquate.
Assignation secrète	Décrire la méthode pour cacher l'assignation des traitements, avec assez de détails pour évaluer si la randomisation peut être prévue à l'avance.	Biais de sélection (allocation des interventions biaisée) en raison d'une assignation secrète inadéquate.
Insu des participants, du personnel et des évaluateurs (à évaluer pour chaque critère de jugement)	Décrire l'ensemble des mesures prises pour que les participants, les soignants et les évaluateurs soient en aveugle du traitement reçu	Biais de performance et de détection dus à la connaissance par les participants et les soignants des interventions allouées.
Données manquantes (à évaluer pour chaque critère de jugement)	Décrire si les données manquantes (perdus de vue ou patients exclus de l'analyse) sont rapportées pour chaque groupe, ainsi que les raisons et nombres.	Biais d'attrition en raison de la quantité, de la nature ou du traitement de données des données manquantes.
Présentation sélective des critères de jugement	Indiquez comment les auteurs ont examiné la possibilité d'une présentation sélective des critères de jugement.	Biais de présentation du fait d'une présentation sélective des critères de jugement
Autres risques de biais	Décrire toutes les autres sources possibles de biais, non évaluées dans les autres domaines	Biais dû à des problèmes non évalués ailleurs dans le tableau.

Tableau 1. « Risk Of Bias tool » (RoB) de la Collaboration Cochrane

Les résultats de l'évaluation du risque de biais peuvent être présentés sous forme tabulaire, contenant pour chaque domaine le jugement des auteurs ainsi que les éléments rapportés dans l'article le justifiant ou sous la forme d'un graphique (fig. 5 et tableau 2). Cette évaluation devra être répétée pour chaque critère de jugement pour les biais d'évaluation, d'attrition et pour le biais de présentation sélective des résultats.

Alvarez-Lerma 2001a

Methods	RCT Sepsis
Participants	140 adult patients hospitalized in the ICU, mechanically ventilated and diagnosed with pneumonia. All infections were hospital acquired. 66% of patients were on inotropic drugs upon entry to study
Interventions	Meropenem 1grX3 for 9.3 days vs. ceftazidime 2grX3 + amikacin 7.5mg/kgX2 for 8.3 days
Outcomes	Overall mortality Treatment failure (clinical and bacteriological) Bacterial superinfections Adverse events Duration of treatment
Notes	Multicentre Spain Outcomes in subgroups: Gram negative and Pseudomonas sp. infections

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer generated in blocks of 6
Allocation concealment (selection bias)	Low risk	Central randomization and by sealed opaque envelopes
Incomplete outcome data (attrition bias) Mortality	Low risk	All 140 randomized patients were evaluated for mortality outcome
Incomplete outcome data (attrition bias) Failure	High risk	116 out of 140 randomized patients were evaluated for failure outcome
Other bias	Unclear risk	Fixed time for outcome assessment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Open
Blinding of outcome assessment (detection bias) All outcomes	High risk	Open

Tableau 2. Présentation tabulaire de l'évaluation du risque de biais d'un essai contrôlé randomisé (Alvarez-Lerma 2001a) inclus dans une revue Cochrane. Tableau extrait de Paul *et al* (136)

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Incomplete outcome data (attrition bias): Mortality	Incomplete outcome data (attrition bias): Failure	Other bias	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)
Abrams 1979	?	?	-	-	?	-	-
Aguilar 1992	?	?	?	+	?	-	-
Alvarez-Lerma 2001 a	+	+	+	-	?	-	-
Wiecek 1986	?	?	+	+	?	-	-
Wing 1998	+	+	+	+	?	-	-
Yellin 1993	+	+	-	-	?	-	+

Figure 5. Exemple de graphique rapportant les résultats de l'évaluation du risque de biais dans une revue Cochrane. Chaque ligne horizontale correspond à un essai inclus dans la revue, chaque carré à l'évaluation du risque de biais de l'un des domaines du RoB tool. La couleur rouge signale un risque de biais élevé, la couleur jaune signale un risque de biais incertain et la couleur verte signale un risque de biais faible. Figure extraite de Paul *et al* (136)

L'utilisation du RoB tool permet de standardiser l'évaluation de la validité interne des ECRs et par extension de les hiérarchiser selon celle-ci. Dans les revues systématiques avec méta-analyses, il est recommandé de restreindre la méta-analyse aux essais à faible risque de biais et/ou de présenter des analyses stratifiées selon le risque de biais des ECRs inclus (32, 138). Or selon l'étude méthodologique publiée par S.Hopewell *et al* seuls 11 % des revues Cochrane et 8 % des revues non-Cochrane prennent en compte les évaluations du risque de biais dans la méta-analyse, et c'est principalement fait à l'aide d'analyses de sensibilité n'incluant que les essais à faible risque de biais, ou bien à l'aide d'analyses en sous-groupes analysant séparément les essais à faible risque de biais et ceux à risque élevé ou incertain (139). La majorité des

revues systématiques incluent dans la méta-analyse l'ensemble des études identifiées et se contentent de décrire le risque de biais dans chacune d'elles. Cette stratégie présente l'avantage de la simplicité. Mais inclure tous les essais cliniques, indépendamment de leurs caractéristiques, dans les méta-analyses fait courir le risque que des décisions médicales soient fondées sur les résultats d'études biaisées. Des biais à l'échelle d'un essai vont impacter les résultats des méta-analyses de manière prolongée dans le temps. Il existe donc une absolue nécessité à choisir les méthodes appropriées lors de la planification et conduite des ECRs afin de minimiser les biais, et ainsi de générer des résultats fiables pour les patients, soignants et chercheurs (140).

Choix des critères de jugement

Le choix des critères de jugement est également une étape fondamentale de la planification des essais cliniques. Les essais cliniques ne sont crédibles que si leurs critères de jugement le sont (141).

Hormis leurs qualités métrologiques intrinsèques (reproductibilité, validité et sensibilité au changement), la pertinence des critères de jugement choisis préjugera de notre capacité à interpréter les effets des interventions testées (142, 143). La réponse quant à l'utilité d'une intervention thérapeutique resterait incomplète s'il n'existait pas des critères de jugement qui permettent de juger l'impact clinique de celle-ci sur les patients. Les chercheurs sont encouragés à estimer les effets des interventions testées sur des critères de jugement importants

pour les patients (143-148). Or cela n'est pas toujours le cas. Il a été rapporté par exemple que moins d'un essai sur cinq dans le domaine du diabète et moins du quart des essais en cardiologie incluaient des critères de jugement importants pour les patients (CJIP) (149, 150). Dans le domaine de la réanimation, une étude récente rapporte que moins d'un quart des critères de jugement étaient des CJIP (24 %) (151) et que cette proportion chutait à 5 % si on excluait la mortalité (e.g. qualité de vie ou fonction) (151). Dans le domaine de la neurologie, le critère de jugement souvent utilisé pour l'évaluation des inhibiteurs de la cholinestérase dans la maladie d'Alzheimer était les échelles de fonction cognitive. Celles-ci permettent de mettre en évidence des différences même infimes des fonctions cognitives, sans que ces celles-ci ne soient cliniquement pertinentes (152). L'implication de patients dans la recherche évaluant les traitements de la polyarthrite rhumatoïde a montré que, pour la plupart des patients, la fatigue était le symptôme dominant alors que les chercheurs se concentraient majoritairement sur la douleur (153).

La quête de la puissance statistique adéquate amène certains à utiliser des critères de jugement cliniquement triviaux et scientifiquement non pertinents (154). Ainsi, au lieu de recourir au CJIP, les chercheurs utilisent des critères de jugement de substitutions («surrogates endpoints»)(155). Ces derniers ont pour grande force de se modifier plus rapidement que le CJIP. Ils permettent ainsi des essais avec des effectifs moindres, et donc des études plus rapides et moins coûteuses (154-157). Ils ne mesurent pas directement un effet thérapeutique important pour le patient, mais des variables, souvent physiologiques, susceptibles de se modifier sous l'effet de la thérapeutique évaluée. Les critères de jugement biologiques, radiologiques ou physiologiques peuvent permettre de comprendre la façon dont une intervention affecterait l'état de santé des patients sans pour autant informer de l'impact clinique de celle-ci. Ainsi, ces

critères de substitution peuvent être trompeurs et retrouver des effets traitement exagérés ou même un bénéfice apparent pour des traitements nocifs, comme ce fut le cas pour l'utilisation des anti arythmiques (Encainide) dans la prévention des extrasystoles ventriculaires après un infarctus du myocarde qui entraînent une diminution du taux d'extrasystoles, mais une augmentation de la mortalité (158-161).

Hormis la pertinence et l'importance clinique des critères de jugement, les utilisateurs de la recherche attendent que les ECRs évaluant les mêmes questions de recherche utilisent et rapportent les mêmes critères de jugement, afin de pouvoir comparer les études entre elles. Certains critères de jugement sont d'une telle importance que tous les essais, et par extension toutes les revues systématiques devraient les mesurer et les rapporter. Mais il existe souvent une forte hétérogénéité dans les critères de jugement retrouvés (162-166). Saldanha *et al*, dans une étude portant sur des essais dans le VIH/SIDA, rapportent que 70 % à 80 % des critères de jugement des essais individuels n'étaient pas inclus dans les revues systématiques évaluant les mêmes interventions, et qu'une grande partie de ces critères de jugement n'étaient rapportés qu'une seule fois (166). Dans une revue méthodologique en pédiatrie, sur 138 essais cliniques ayant mesuré la diarrhée aiguë, 64 ont utilisé des définitions uniques de la diarrhée, 69 des définitions uniques de la résolution de celle-ci et plus du tiers utilisaient des critères de jugement retrouvés dans aucun autre essai (165). L'absence de consensus sur le ou les critères de jugement à mesurer entraîne une hétérogénéité importante entre les essais et ainsi limite fortement les possibilités de synthèse des éléments de preuve (167). Si la recherche ne répond pas aux besoins de santé et aux priorités de ses usagers, si elle n'est pas synthétisable, les données produites seront moins utiles et auront moins d'effet autant en terme clinique que de santé publique. Ainsi, le choix des critères de jugement dans les essais cliniques pourrait nuire

à la capacité des usagers de la recherche à prendre des décisions éclairées (22, 73-77, 79, 168-170).

1.2.3 Accessibilité des résultats de la recherche

Non-publication des résultats d'un essai clinique

Une fois un essai clinique mené, celui-ci devra être publié pour que ses conclusions puissent bénéficier aux usagers de la recherche. C'est par le biais d'articles de recherche, publiés dans des revues scientifiques, que les résultats des essais cliniques sont communiqués à la communauté scientifique. La publication, ou non-publication des résultats ne devra pas être dépendante de la magnitude et de la direction de l'effet thérapeutique éventuel. Ainsi un essai non publié ne pourra contribuer au corpus de connaissances existantes et ne pourra pas aider les usagers de la recherche. La non-publication des résultats d'un essai clinique est inacceptable autant sur le plan scientifique que sur le plan éthique (171). Celle-ci rompt le contrat implicite liant les patients et les investigateurs. Bien que ce problème soit connu depuis de nombreuses années près de la moitié, des essais cliniques ne sont pas publiés (fig. 6) (172-177). Cette proportion est stable depuis près de 30 ans (172-176).

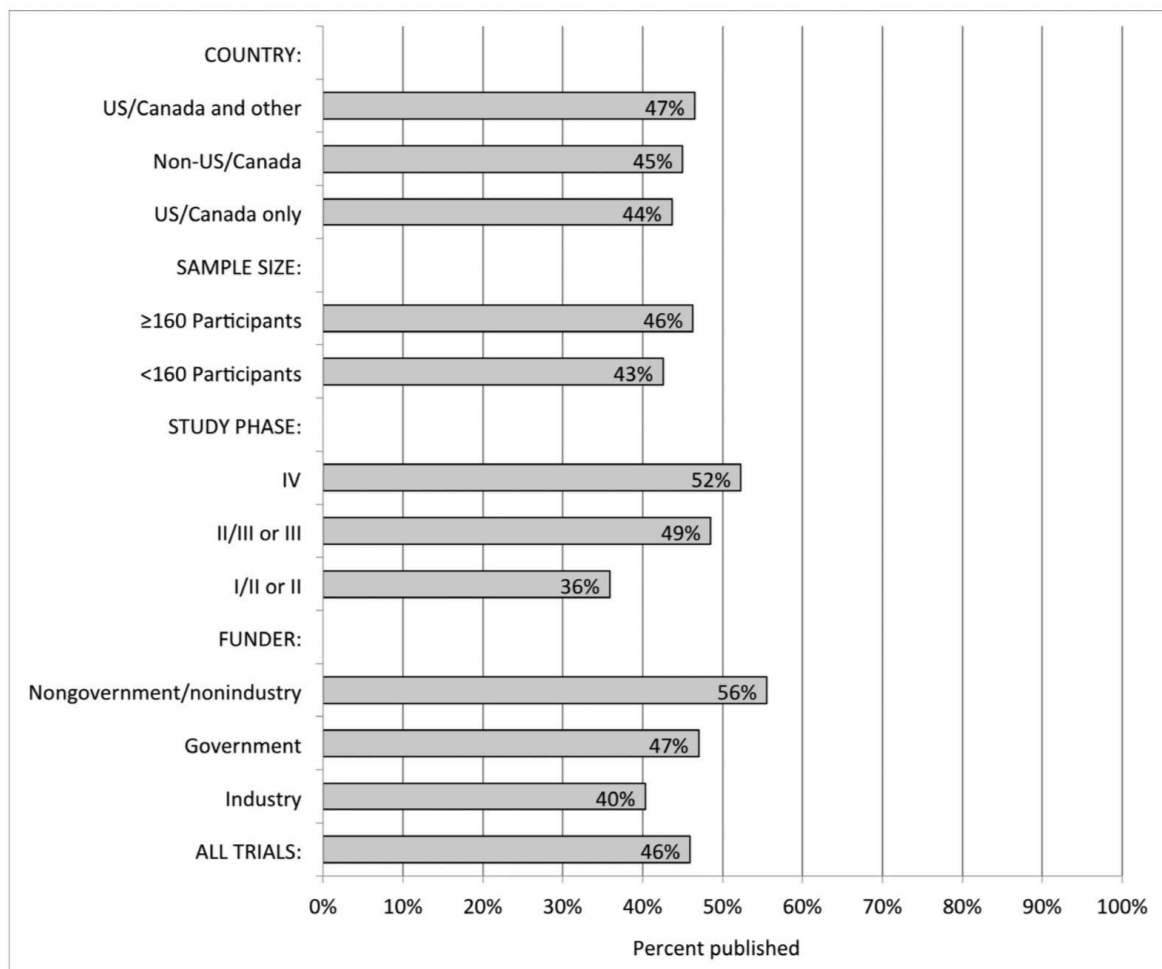


Figure 6. Taux de publication d'un échantillon aléatoire de 677 essais cliniques enregistrés sur ClinicalTrials.gov de 2000 à 2007. Figure extraite de Chan *et al* (75) et basée sur les données de Ross J *et al* (178).

Par exemple, dans une étude ayant analysé 3668 projets de recherche financés par la Commission européenne à hauteur de 3780 millions d'euros sur une période de 9 ans, des chercheurs ont établi que respectivement 49,9 % et 44 % des projets avaient donné lieu à une publication indexée dans Google Scholar ou PubMed (176). Ce qui correspondrait à 570 millions d'euros d'investissements sans résultats académiques facilement détectables pour la communauté scientifique. Au cours d'une revue systématique portant sur l'efficacité du traitement péri opératoire par prégabaline dans la prévention de la douleur chronique, des

chercheurs ont décrit que 60 % des essais inclus dans la revue étaient non publiés, ce qui correspondait à 79 % des patients inclus (fig. 7) (179).

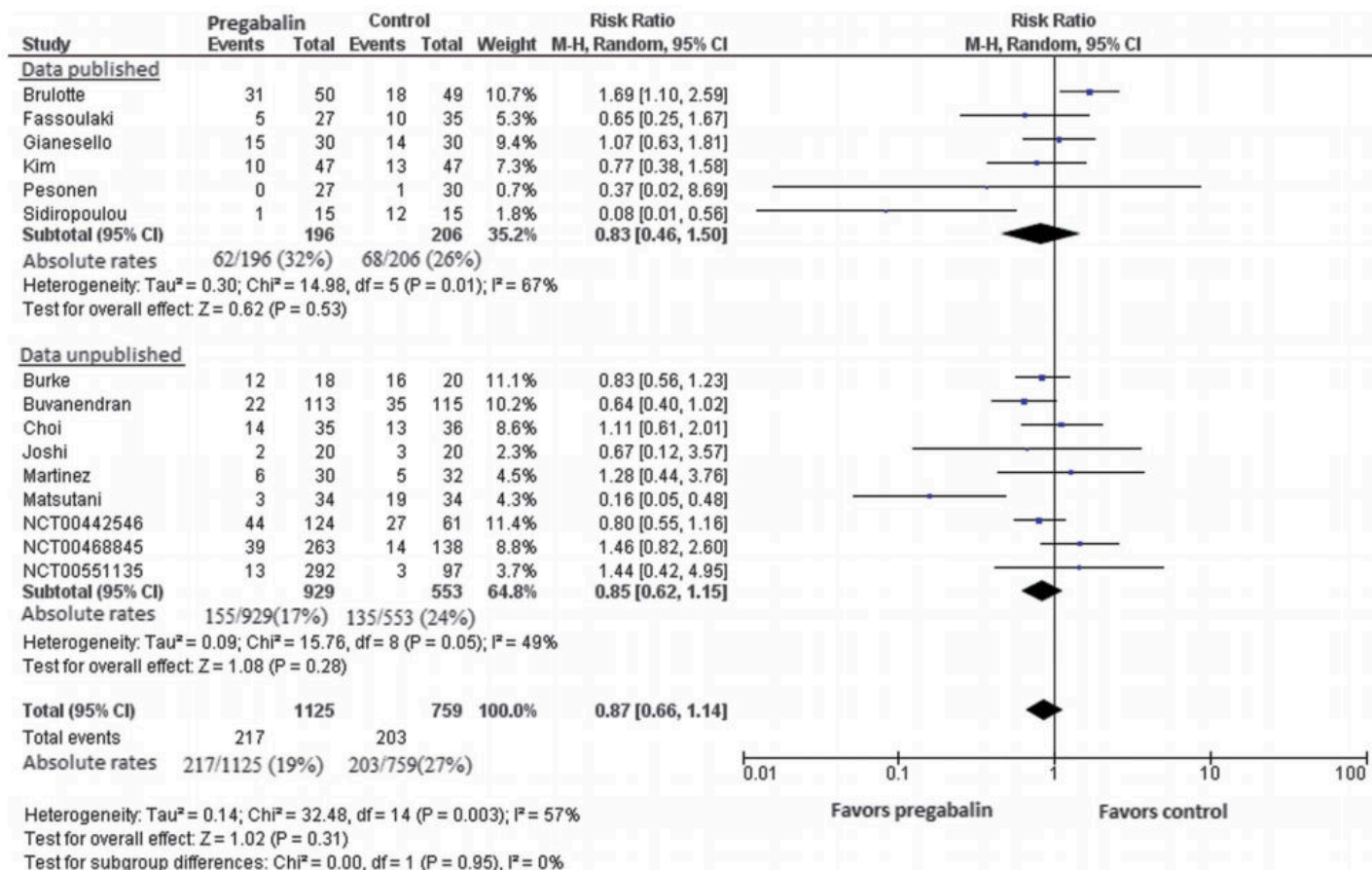


Figure 7. Diagramme en forêt illustrant le problème des essais non publiés dans le domaine de la prise en charge de la douleur chronique. Figure extraite de Martinez *et al* (179).

Dans une autre étude, des chercheurs ont évalué l'ensemble des essais soumis à la FDA (Food and Drug Administration) pour l'autorisation de mise sur le marché de 12 antidépresseurs entre 1987 et 2004 aux États-Unis (41). Parmi les 74 essais retrouvés dans les documents réglementaires, 31 % n'ont jamais été publiés dans une revue scientifique. Trente-sept études considérées par la FDA comme ayant des résultats positifs (c'est-à-dire favorisant le traitement expérimental) avaient été publiées et une étude considérée comme positive n'a pas été publiée.

Les études considérées par la FDA comme ayant des résultats négatifs ou non probants étaient, sauf à 3 exceptions près, non publiées (22 études) ou d'une manière suggérant des résultats positifs (11 études). Si l'on se basait sur les données publiées, 94 % des essais identifiés étaient positifs. En revanche, en tenant compte des seules données de la FDA, 51 % étaient positifs. Ceci avait pour conséquence une surestimation de l'effet traitement des antidépresseurs de l'ordre de 32 % (41). Ce travail de Turner *et al*, illustre parfaitement le biais de publication qui affecte la recherche biomédicale (41). Ce biais correspond à la proportion des chercheurs à publier plus souvent les essais avec des résultats positifs que les essais avec des résultats ne retrouvant pas de différence statistiquement significative (45, 175). Mais ce biais de publication est possiblement plus discret quand il ne s'agit plus de non-publication, mais de publication avec des délais plus longs entre l'achèvement et la publication en fonction de la direction et de la significativité des résultats (175, 180-182). Dans un travail ayant analysé 61 études sur les effets sur la santé du tabagisme passif, Misakan *et al* décrivent que le temps médian depuis le début du financement à la publication était de 5 ans pour les études statistiquement non significatives et 3 ans pour les études significatives ($p = 0,004$) (181). En oncologie, les travaux publiés sous forme d'abstracts aux congrès de l'American Society of Clinical Oncology (ASCO) sur près de 10 ans de conférences, et présentant des résultats statistiquement significatifs étaient publiés plus rapidement que ceux avec des résultats négatifs (2,2 vs 3 ans de délai médian)(182).

L'un des cas les plus emblématiques du biais de publication reste celui-ci des antiviraux, et en particulier l'oseltamivir (Tamiflu). En 2009 la Collaboration Cochrane publie une revue systématique de la littérature sur l'efficacité des inhibiteurs de la neuraminidase dans la prévention et le traitement de la grippe chez les adultes en bonne santé (183). Cette revue a mis

en évidence que huit des 10 ECR n'avaient jamais été publiés. Et que les deux seuls qui avaient été publiés étaient financés par Roche, et rédigés par des employés de cette même firme. Ceci avait amené la Collaboration Cochrane, avec l'aide du BMJ (British Medical Journal), à réclamer à de nombreuses reprises les données manquantes (184, 185). Celles-ci ont révélé que les données de 60 % des patients inclus n'avaient jamais été publiées. L'analyse de l'ensemble des données remettait fortement en cause le rapport bénéfices-risques de l'oseltamivir. La conséquence directe de ces données partiellement accessibles a été une surestimation de l'effet traitement, ayant poussé les gouvernements à des dépenses de plusieurs milliards de dollars afin de constituer des stocks dans le monde entier pour un médicament dont les bénéfices en termes de réductions des hospitalisations et des complications pulmonaires dues à une pandémie de grippe sont très discutables (186). Le développement des registres d'essais cliniques tel que clinicaltrials.gov ou clinicaltrialsregister.eu, l'obligation d'y enregistrer les essais, mais également d'y publier les résultats doit permettre de limiter le biais de publication (187, 188). Mais, sur un échantillon de 209 essais enregistrés sur clinicaltrials.gov, dans le domaine de l'oncologie, 9 %, 25 % et 36 % avaient posté des résultats à respectivement 12, 24 et 36 mois de la fin des études (189). Cette étude semble indiquer que les registres d'essais permettent d'identifier les essais existants (ou en cours), mais que les résultats n'y sont que rarement rapportés pour le moment. Leurs effets sur le biais de publication semblent pour le moment peu clairs. Des résultats similaires ont été décrits récemment par Goldacre *et al*, sur une cohorte d'essais cliniques enregistrés sur le registre européen d'essais cliniques EUCTR (EU Clinical Trials Register) (190).

1.2.4 Rapports de recherche transparents, non biaisés et complets

Présentation sélective ou incomplète des résultats

Pour pouvoir être utile aux praticiens et aux patients, la recherche publiée doit l'être de manière honnête, complète et transparente. Ils doivent pouvoir avoir confiance dans le fait que les résultats présentés sont justes et non le reflet d'une publication sélective de ceux-ci en fonction de la direction et importance de l'effet. Les auteurs des essais cliniques doivent donc rapporter leurs résultats avec suffisamment de précision pour permettre leur inclusion dans des revues systématiques et méta-analyses dont le but est d'obtenir une synthèse exhaustive des données disponibles.

En lisant la partie résultats d'un article, le lecteur s'attend à y retrouver une transcription fidèle des effets de l'intervention testée. Les critères de jugement et les analyses devraient correspondre à ce qui était prévu dans le protocole. Malheureusement, les rapports d'essais cliniques ne sont pas toujours honnêtes et peuvent souffrir du biais de présentation sélective des résultats et des critères de jugement. Celui-ci est observé quand la présence d'un critère de jugement dans le rapport d'un essai semble liée à la nature et à la direction des résultats. Les cas de présentations sélectives de résultats ont été regroupés par Page *et al* (191) dans diverses catégories telles que l'omission sélective des résultats pour certains critères, la mention sélective parmi de nombreuses mesures d'un même critère (à des moments différents ou présentation d'un score global versus un sous-score, etc.), la présentation sélective d'une transformation des résultats (variables continues transformées en variables dichotomiques), présentation sélective parmi de multiples modèles d'analyse statistique, présentation sélective

de certaines sous-populations, ou encore présentation sélective des résultats dans les différentes parties d'un même rapport ou dans différents rapports pour un même ECR. De nombreux travaux ont exploré ces biais en se penchant sur les discordances entre les critères de jugement présentés dans les protocoles et ceux retrouvés dans les rapports des essais cliniques (57, 59, 192-195). Les discordances observées incluent l'omission du critère de jugement principal préspecifié, l'introduction d'un nouveau critère de jugement principal, des critères de jugement principaux devenant secondaires ou inversement et des périodes de mesure des critères de jugement qui différaient (57, 193). Plus d'un tiers des critères de jugement planifiés seraient présentés sélectivement sur la base de la nature et de la direction des résultats observés (57, 59, 192-196). Dans leurs travaux, Mathieu *et al*, rapportent par exemple que parmi les articles dont les essais ont été correctement enregistrés, 31 % (46 sur 147) présentaient des discordances entre les critères de jugement planifiés et publiés. L'analyse de la moitié de cas discordants a permis de mettre en évidence que dans plus de 80 % des cas ces discordances apparaissaient au profit de résultats statistiquement significatifs (193).

Hormis la présentation sélective, la présentation incomplète des critères de jugement est une autre menace qui pèse sur l'utilité des rapports d'essais cliniques. En 2004, Chan *et al* décrivaient que près de 50 % des critères de jugement d'efficacité et 65 % des critères de sécurité sont décrits de manière incomplète dans les rapports d'ECR (57). De 30 à 50 % des critères de jugement seraient rapportés de manière incomplète (57, 197, 198). Ces descriptions partielles rendent leur inclusion dans les méta-analyses impossibles et conduiraient ainsi à une vision tronquée des preuves disponibles, en particulier sur les données relatives aux effets indésirables. Dans une revue systématique avec une méta-analyse en réseau Martinez *et al* (199)

rapportent par exemple que sur 60 essais ayant comparé les anti-inflammatoires non stéroïdiens à un placebo, 15 rapportaient les effets indésirables (fig. 8).

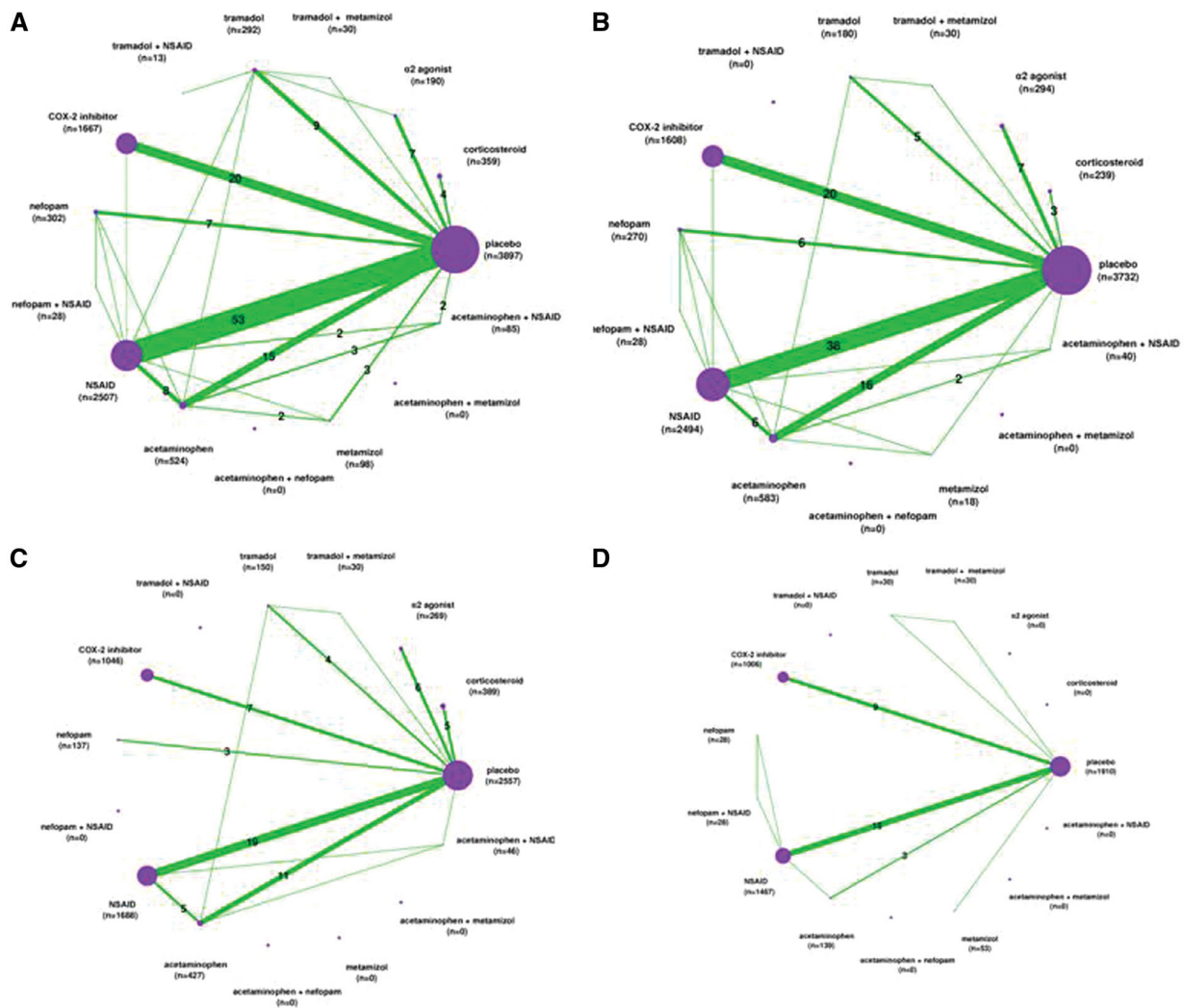


Figure 8. Géométrie du réseau pour les essais rapportant un effet du traitement pour A) douleur, B) nausée, C) vomissement et D) effets indésirables dans une méta-analyse en réseau portant sur les analgésiques non opiacés dans la douleur postopératoire. Figure extraite de Martinez *et al* (199).

Une autre étude a récemment analysé les essais cliniques évaluant les probiotiques, prébiotiques et/ou symbiotiques (200). Sur les 384 essais inclus, aucune donnée sur la sécurité des interventions n'était rapportée dans 106 essais (28 %). Or un manque de données sur les effets indésirables de toute intervention thérapeutique peut avoir des conséquences dramatiques sur les patients.

Ces dernières années ont été émaillées de scandales sanitaires impliquant diverses molécules telles que le benfluorex ou le rofécoxib (201-203). Dans le cas du rofécoxib, la principale étude avait rapporté des données présentant des dates d'interruption différentes pour les événements cardiovasculaires et gastro-intestinaux. Cette présentation trompeuse des résultats des effets indésirables n'a pas été pleinement décrite dans la publication et a eu pour effet de favoriser l'effet du médicament sur les événements gastro-intestinaux tout en minimisant le risque d'événements cardiovasculaires (202, 203).

Exhaustivité des rapports de recherche

De nombreuses études ont décrit la lenteur de diffusion des connaissances issues de la recherche à la pratique médicale courante. Cette durée a souvent été estimée à près de 17 ans (204-207). Parmi les raisons rapportées pour expliquer cet important délai, ont été évoquées l'ignorance de l'existence de nouvelles données, la perception que ces résultats sont non valides ou enfin que ceux-ci sont non applicables au contexte, et patients, des cliniciens (208, 209). La

digitalisation et la généralisation d'outils tels que la « Free Open Access Medical education » (FOAMed) ou le « post publication peer review » pourraient aider à réduire ce délai (210-213).

Mais l'obstacle majeur à la dissémination et à l'utilisation des nouvelles connaissances est la qualité du rapport de recherche. Sans un rapport de recherche précis et complet, les soignants ne peuvent faire profiter les patients des bénéfices thérapeutiques mis en évidence dans les essais cliniques (214-217). Une description exhaustive des interventions médicamenteuses et non médicamenteuses évaluées dans les essais cliniques est un préalable essentiel à leur mise en application dans la pratique clinique. Or celles-ci semblent encore insuffisamment décrites, avec des descriptions insuffisantes et partielles dans près de 40 % des essais cliniques (215). Cette proportion serait même de l'ordre de 80 % dans les essais évaluant des interventions complexes, telles que des interventions d'éducation thérapeutique (214). Mais la proportion d'interventions reproductibles grimperait à 90 % après obtention d'informations supplémentaires auprès des auteurs (fig. 9) (215), laissant ainsi supposer qu'il s'agirait essentiellement d'un problème lié à une description incomplète dans l'article initial.

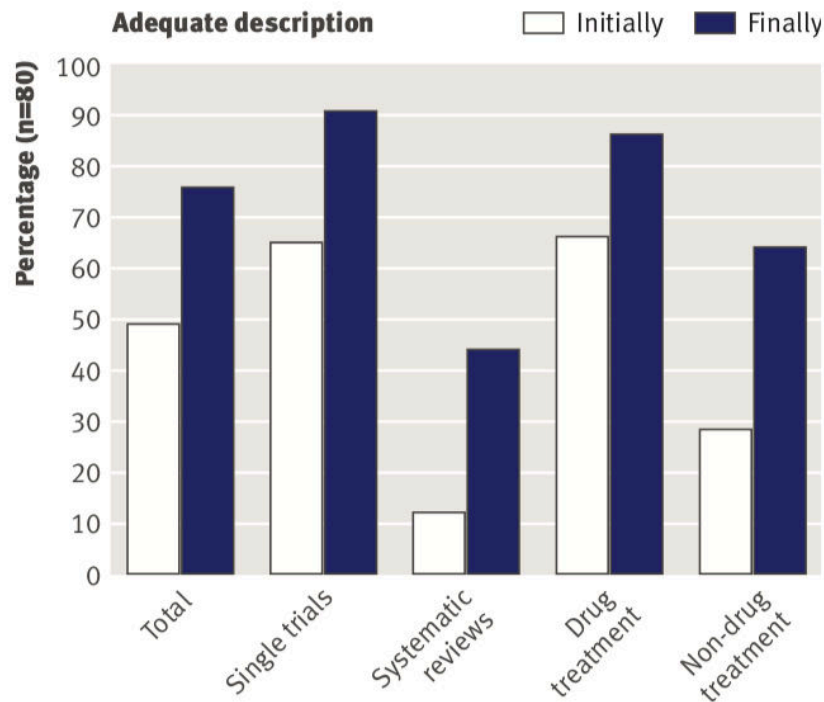


Figure 9. Pourcentage d'études présentant une description suffisante de l'intervention initialement (sur la seule base du document publié) et après l'obtention d'informations supplémentaires auprès des auteurs. Figure extraite de Glasziou *et al* (215).

La complexité de certaines interventions pourrait expliquer les détails manquants et de nombreux auteurs semblent disposés à les fournir pour permettre la reproduction de celles-ci (215). Ainsi, par essence, un problème de mise en forme de la recherche devrait donc pouvoir être résolu par une meilleure description des interventions dans l'article ou par la mise à disposition de matériels supplémentaires. Ces derniers peuvent avoir la forme d'appendices à l'article original ou de manière plus rare, mais permettant ainsi une plus grande liberté de format, sous la forme d'un site internet comme proposé par les auteurs des essais CRASH (218-221). Chacun de leurs essais déjà publiés, ou en cours, sont accompagnés d'un site internet avec de nombreux documents décrivant l'ensemble des détails et permettant de reproduire les interventions testées (moment du traitement, dose, durée d'administration, modalités de surveillance, matériel, instructions, etc.)

Embellissement des résultats

Les auteurs d'essais cliniques peuvent utiliser des stratégies afin d'embellir leurs données et ainsi être à l'origine d'une interprétation biaisée de ceux-ci. Le but de telles manœuvres est de mettre en avant un bénéfice de l'intervention testée, sans que cette affirmation ne soit supportée par les résultats de l'essai. Ces stratégies ont été regroupées sous le terme de « spin ». Ces exagérations seraient dues à une méconnaissance des chercheurs des bonnes règles de rédactions des rapports d'ECR ou encore à un enthousiasme exagéré de ces derniers. Ils peuvent céder à la tentation d'amplifier leurs résultats. Dans une analyse de 72 essais contrôlés randomisés dits négatifs (n'ayant pas retrouvé de différence significative pour le critère de jugement principal), Boutron *et al* ont rapporté que près de 30 %, 40 % et 50 % des sections résultats, discussions et conclusions des articles présentaient une forme d'embellissement des résultats ou spin (222). L'analyse a mis en évidence un certain nombre de stratégies de spin, les plus courantes consistaient par exemple à focaliser l'attention sur une analyse en sous-groupe ou sur des résultats de critères de jugement secondaires statistiquement significatifs ou enfin à affirmer l'équivalence des traitements quand les résultats n'étaient pas statistiquement significatifs. D'autres auteurs ont analysé des essais dans le domaine de l'oncologie et retrouvaient dans près de la moitié des essais une forme de spin dans les résumés ou dans le texte de l'article, et que la présence de ces spins avait un impact sur l'interprétation des résultats par des lecteurs (223, 224).

1.3 Justification de la démarche de thèse

La dénonciation de la recherche clinique de mauvaise qualité n'est pas récente. Le constat est fait depuis des années ; de nombreuses études sont menées pour les mauvaises raisons, mal conçues, mal exécutées, non rapportées, insuffisamment rapportées ou encore rapportées de manière déformée. La littérature biomédicale représenterait ainsi un sous-ensemble incomplet et biaisé de la recherche existante. L'ensemble de ces conduites ont été rassemblées sous le label commun de « gaspillage de la recherche ». Chalmers et Glasziou ont tenté de quantifier les « énormes sommes d'argent » évoquées par Douglas Altman dans son réquisitoire contre la mauvaise recherche médicale (22, 65). Ils ont ainsi estimé que 85 % des plus de 100 milliards de dollars dépensés annuellement au niveau mondial pour la recherche médicale seraient gaspillés (22). Ce gaspillage n'a pas que des conséquences financières. Il nuit à la mise en œuvre, et en pratique, des conclusions de la recherche. Il compromet la capacité des professionnels de santé à prodiguer des soins appropriés, efficaces et basés sur les données de la recherche. Une telle situation est indéfendable sur le plan scientifique, économique et éthiques (225, 226). Dans la série publiée par le Lancet (74-77), les auteurs reviennent sur les conséquences de celui-ci, sur les éventuelles solutions afin de le réduire et ainsi d'augmenter la valeur de la recherche. Dans le cadre de cette thèse, nous faisons l'hypothèse qu'une part du gaspillage de la recherche aurait pu être évitée à l'aide de moyens simples et peu coûteux. Nous nous sommes plus spécifiquement intéressés au gaspillage de la recherche lié à une mauvaise planification des ECRs (choix des méthodes et des critères de jugement) ou à présentation sélective ou incomplète des critères de jugement.

Dans un premier temps nous nous sommes spécialement intéressés au gaspillage évitable de la recherche lié à l'emploi de méthodes inappropriées dans des essais cliniques inclus dans des revues de la Collaboration Cochrane. Partant du présupposé qu'un essai clinique biaisé ne pouvait pleinement participer aux corpus de connaissances, nous avons cherché à estimer la part et causes des biais qui affectent un panel d'ECR évaluant une intervention thérapeutique. Puis nous avons ensuite estimé dans quelle mesure ce gaspillage pouvait être évité de manière simple et peu coûteuse.

Dans un second temps, nous nous sommes intéressés au gaspillage de la recherche lié aux critères de jugement des essais cliniques, ceci impliquant de s'intéresser à la planification des ECRs (choix des critères de jugement) ainsi qu'à la manière dont les critères de jugement étaient rapportés (présentations sélectives ou incomplètes des critères de jugement). Afin de pouvoir apprécier celle-ci, nous avons d'abord cherché à identifier les critères de jugement les plus importants à mesurer dans le cadre de l'évaluation d'une thérapeutique. Puis nous avons estimé la part des ECRs pour lesquels ces derniers étaient manquants, rapportés de manière sélective ou partielle. Puis nous avons estimé dans quelle mesure les critères de jugement manquants auraient pu être mesurés de manière simple et peu coûteuse.

Pour chaque travail, nous résumerons les méthodes ainsi que les principaux résultats avant de présenter les tirés-à-part des publications. Enfin, le chapitre 5 de cette thèse sera consacré à la discussion de nos résultats et les perspectives de nos travaux de recherche.

2. Gaspillage évitable de la recherche dû à l'emploi de méthodes inadéquates dans les essais cliniques

2.1. Justification de l'étude

Des défauts dans la conception, la conduite et l'analyse des essais cliniques peuvent biaiser les résultats. À l'aide de travaux métaépidémiologiques, de nombreux chercheurs ont identifié et décrit les caractéristiques des ECRs associées à l'effet traitement (119-135). Aucune étude n'a évalué la part du gaspillage de la recherche qui aurait pu être évitée si les chercheurs avaient eu recours à des méthodes adéquates.

2.2. Objectifs

L'objectif de ce travail était d'évaluer le gaspillage de la recherche dû à des méthodes inadéquates dans les ECRs et d'examiner dans quelle mesure ce gaspillage pourrait être évité.

2.3. Méthodes

Nous avons mené une revue méthodologique, suivie d'une étude de simulation, d'essais cliniques inclus dans des revues systématiques Cochrane publiées récemment.

Sélection des revues Cochrane, des méta-analyses et des essais

Nous avons recherché au sein de la « Cochrane Database of Systematic Reviews » toutes les revues systématiques Cochrane :

- a. publiées entre le 1er avril 2012 et le 31 mars 2013
- b. ayant inclus des ECRs avec au moins une méta-analyse
- c. ayant examiné les effets d'une intervention thérapeutique
- d. ayant utilisé l'outil d'évaluation du risque de biais de la collaboration Cochrane : le Risk of Bias tool.

Pour chaque revue incluse, nous avons sélectionné la méta-analyse du critère de jugement principal tel que défini par les auteurs de la revue. Si plusieurs critères de jugement principaux étaient rapportés, nous avons sélectionné la méta-analyse ayant inclus le plus grand nombre d'essais. Enfin, nous avons examiné tous les ECRs inclus dans les méta-analyses identifiées par la méthode décrite ci-dessus.

Évaluation du risque de biais

Afin de déterminer si les essais cliniques utilisaient des méthodes inadéquates, nous nous sommes basées sur l'évaluation de leur risque de biais. La collaboration Cochrane recommande de systématiquement évaluer le risque de biais des ECRs inclus en utilisant le RoB tool et non des échelles de qualité (138, 227). Nous avons ainsi extrait l'évaluation du risque de biais, tel que rapporté par les auteurs des revues, pour les domaines suivants : génération de la séquence

de randomisation, assignation secrète, insu des participants, du personnel et des évaluateurs et enfin les données manquantes. Puis pour un échantillon aléatoire de 200 ECRs, nous avons, en double et de manière indépendante, réévalué le risque de biais (CR et YY) en nous basant sur les articles des ECRs de notre échantillon. L'évaluation du risque de biais lié à l'insu (biais de performance et de détection) et aux données manquantes a été faite pour chaque critère de jugement (138), en tenant compte de l'aspect objectif ou subjectif de ceux-ci selon Savovic *et al* (128). Nous avons considéré comme objectifs : la mortalité, les autres critères évalués objectivement (par exemple : grossesse, naissances ou résultats de laboratoire), ou encore les critères mesurés objectivement, mais potentiellement influencés par le jugement des cliniciens ou des patients (par exemple : hospitalisations, abandons, accouchement déclenché ou traitements supplémentaires administrés). Les critères de jugement évalués par les cliniciens ou ceux rapportés par les patients ont été considérés comme subjectifs.

Identification des problèmes méthodologiques et ajustements éventuels

Pour tout essai confirmé comme à risque de biais élevé, deux d'entre nous (AD et YY) ont indépendamment identifié les problèmes méthodologiques responsables du risque de biais élevé et ont suggéré des ajustements méthodologiques pour les éviter. Ces ajustements sont basés sur la littérature méthodologique et les règles de bonne pratique en termes de recherche clinique (127, 130, 228, 229). Ensuite, un panel d'experts a évalué les ajustements méthodologiques proposés. Les experts, 4 méthodologistes internationalement reconnus (DA, IB, PR, SH), ont évalué les problèmes méthodologiques et les ajustements suggérés. Ils ont eu à estimer si ceux-ci étaient « simples, modérément simples, difficiles ou impossibles » à implémenter et à quel coût (aucun coût, coût mineur, coût modéré ou coût majeur). L'évaluation finale de la faisabilité

et du coût était basée sur le consensus du groupe d'expert. Pour les traitements pharmacologiques, nous avons contacté le pharmacien en charge des services pharmaceutiques pour tous les essais cliniques pharmacologiques menés au sein de l'Assistance Publique-Hôpitaux de Paris. Pour les traitements non pharmacologiques, nous avons contacté deux chirurgiens (un chirurgien orthopédique et un obstétricien) impliqués dans des essais cliniques. Ces trois experts ont évalué les propositions en aveugle du consensus des experts.

Évaluation du gaspillage évitable de la recherche dû à des méthodes inadéquates

Enfin, nous avons estimé la proportion de gaspillage évitable de la recherche en définissant le gaspillage évitable comme tout ECR dont le risque de biais global passerait d'élevé à faible à l'aide d'ajustements simples et peu coûteux. Nous avons également évalué le gaspillage évitable si tous les ajustements, quel que soit leur coût et difficulté, étaient appliqués. Nous avons enfin réalisé une étude de simulation (appendice 2, tiré-à-part du présent article (168)), afin de réévaluer le gaspillage évitable sur notre échantillon d'ECR, si tous les essais avaient été correctement rapportés. Les analyses statistiques ont été réalisées en utilisant le logiciel R (V 3.0.2. <http://www.r-project.org>).

2.4. Résultats

Nous avons sélectionné 205 revues Cochrane, ayant inclus 1286 ECRs dans la méta-analyse du critère de jugement principal (fig. 10). Les spécialités médicales les plus représentées étaient la psychiatrie (18 %), la gynécologie obstétrique (10 %) et l'oncologie (8 %). Les ECRs inclus ont été publiés entre 1957 et 2012 (année médiane de publication 2004). Le critère de jugement

principal a été considéré comme objectif dans 39 % des essais (n=500), et il s'agissait de la mortalité dans 23 % des ECRs (n=302).

Dans l'échantillon, 16 % des essais (n=207) avaient tous les domaines du RoB tool à faible risque de biais, 43 % (n=556) avaient au moins un domaine à risque élevé et enfin 41 % (n=523) avec au moins un domaine à risque incertain (fig. 10.).

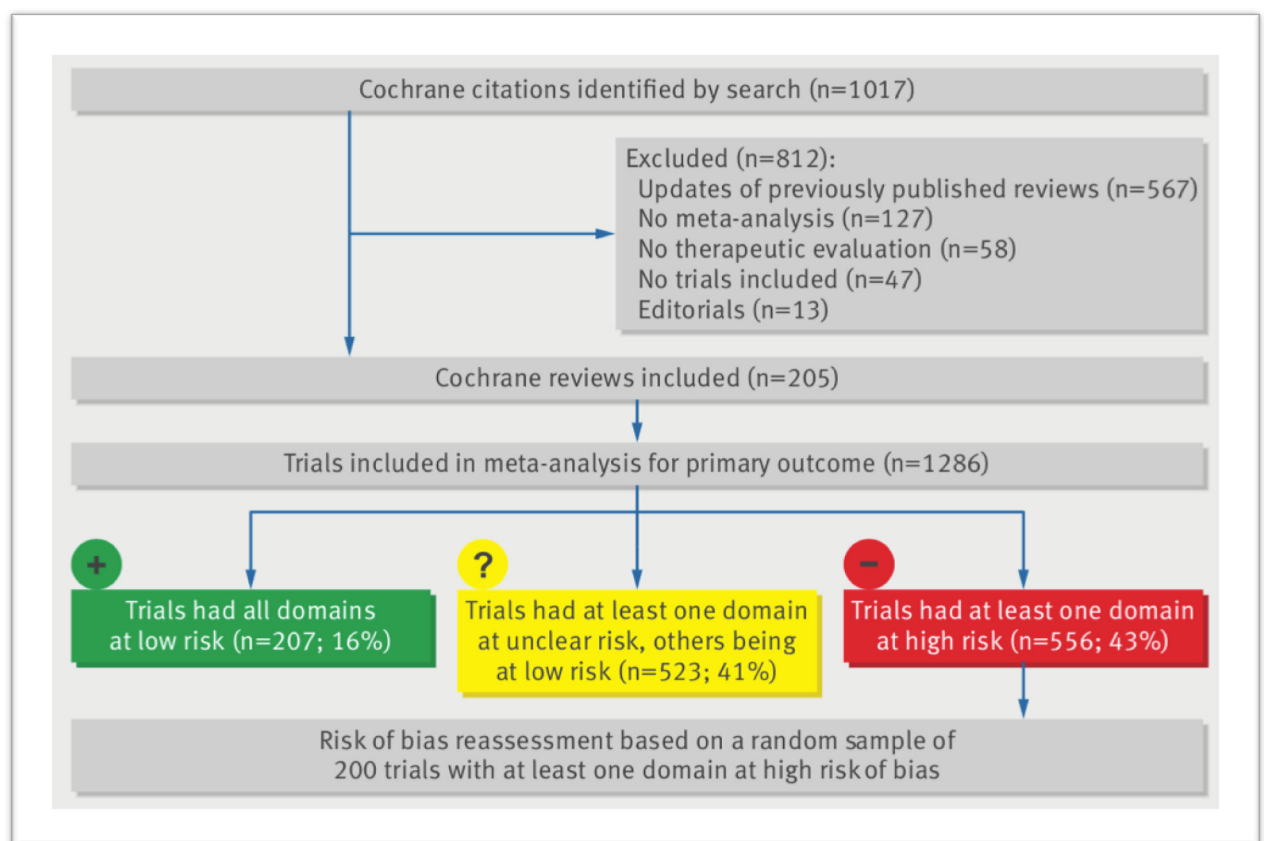


Figure 10. Sélection des revues et essais cliniques, et évaluation du risque de biais.

Parmi l'échantillon aléatoire de 200 essais à risque de biais élevé selon les auteurs, nous avons confirmé celui-ci pour 71 % d'entre eux (n=142). Les désaccords avec les évaluations des auteurs concernaient principalement le biais de détection (pas d'insu, mais critère de jugement

objectif). Nous avons identifié 25 types différents de problèmes méthodologiques parmi ces 142 ECRs. Les plus fréquemment rencontrés étaient l'exclusion des patients de l'analyse (n=50, 35 %), l'absence d'insu avec un critère de jugement rapporté par le patient (n=27, 19 %), l'absence d'insu lors de l'évaluation d'interventions non pharmacologiques (n=23, 16 %), et enfin une méthode inadéquate pour traiter les données manquantes (n=22, 15 %).

Les 25 types de problèmes méthodologiques, les ajustements proposés, leur faisabilité et les coûts selon les experts sont détaillés dans le tableau 3 du tiré-à-part de cet article (168). D'après l'évaluation des experts, au moins un ajustement aurait pu être effectué dans 96 % des ECRs (n=136). L'application d'ajustements méthodologiques simples et peu coûteux aurait pu réduire le nombre de domaines à risque de biais élevé dans 71 essais (50 %). Ces ajustements permettraient à 17 essais (12 %) d'avoir l'ensemble des domaines du RoB à faible risque de biais. Ainsi, le gaspillage évitable de la recherche liés à des méthodes inadéquates représentait 12 % (IC 95 % 7 %-18 %) des essais. Avec tous les ajustements considérés comme possibles par les experts, quels que soient la difficulté ou le coût, le gaspillage évitable de la recherche représentait 31 % (n = 44) (IC 95 % 23 %-39 %) des essais (fig. 11). (IC 95 % 36 % - 49 %).

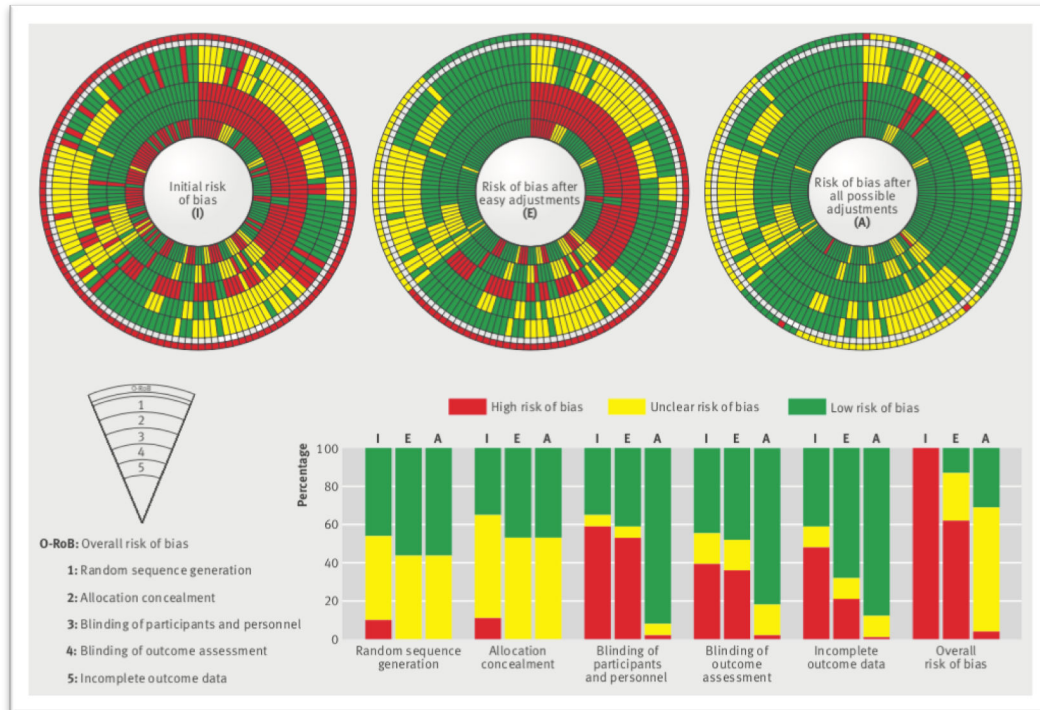


Figure 12. Gaspillage évitable de la recherche liée aux méthodes inadéquates dans les essais avec au moins un domaine confirmé à risque élevé de biais ($n = 142$). Cette figure résume l'évolution du risque de biais par domaine pour chaque essai, initialement (I), après les ajustements méthodologiques simples (E), ou après tous les ajustements méthodologiques (A). Chaque rayon représente l'un des 142 essais à haut risque de biais. Les briques sont une représentation visuelle de l'évaluation du risque de biais par domaine, le rouge indiquant un risque élevé de biais ; le jaune un risque incertain ; et le vert un risque faible. Chaque cercle concentrique représente l'un des domaines du RoB tool, la « génération la séquence de randomisation » étant le cercle le plus éloigné du centre. Le cercle le plus externe représente un aperçu rapide du risque de biais pour chaque essai (rouge si au moins un domaine est à risque élevé, jaune si au moins un domaine est à risque incertain et vert si tous les domaines sont à faible risque). L'histogramme est une représentation du risque de biais, par domaine, parmi les 142 essais à haut risque, initialement (I), après des ajustements méthodologiques simples (E), ou après tous les d'éventuels ajustements méthodologiques (A)

Après la prise en compte des essais insuffisamment rapportés par l'imputation des domaines incertains, 43 % (IC 95 % 39 % - 47 %) des ECRs seraient à faible risque de biais et 57 % (IC 95 % 53 % - 61 %) aurait au moins un domaine à risque de biais élevé. L'application des ajustements méthodologiques simples et peu coûteux permettrait, dans ce cas-là, à 42 % (IC

95 % 36 % - 49 %) des essais de passer de risque élevé à faible, soit un gaspillage évitable de la recherche lié à des méthodes inadéquates estimé à 42 %.

2.5. Discussion

Nos résultats suggèrent qu'appliquer des ajustements simples et peu coûteux permettraient de limiter le risque de biais dans 50 % des essais, et ainsi réduire partiellement le gaspillage de la recherche. Ces résultats peuvent s'expliquer par l'absence fréquente d'implication des méthodologistes et des statisticiens au stade de la planification des essais, mais aussi par une méconnaissance des chercheurs des bonnes pratiques en termes de méthodes de recherche (77, 230). Une formation spécifique aux méthodes de recherche ou à la lecture critique de celle-ci semble cruciale. Tous les médecins ne seront pas des chercheurs, mais tous doivent être capables de lire de manière critique un article afin de guider leurs décisions médicales. Parmi notre large échantillon d'essais, 43 % étaient à risque de biais élevé et 41 % étaient classés incertains, largement en raison de rapports incomplets. Ce chiffre est proche des données de la littérature (231-233) et confirme le problème important du aux rapports imparfaits. Bien qu'un rapport incomplet ne signifie pas nécessairement que l'étude ait été mal planifiée et menée (234, 235), cela reste une autre cause importante de gaspillage de la recherche, l'essai ne pouvant alors pas être correctement évalué par le lecteur(76). Le gaspillage évitable de la recherche représenterait 12 % des ECRs de notre échantillon. Mais si les rapports étaient complets, cette part pourrait représenter jusqu'à 42 % des essais. D'importants efforts ont été faits pour améliorer les rapports de recherche en développant et diffusant largement des recommandations sur la meilleure façon de produire un rapport d'études (25, 236). Ces recommandations existent pour différents types d'études et sont promues par le réseau EQUATOR (Enhancing the QUALity and Transparency Of health Research) (237). Il s'agit d'une initiative internationale

qui tient à jour une liste des recommandations conseillées ou requises dans le cadre de la publication scientifique dans le domaine de la santé. Mais l'utilisation de ces recommandations reste pour le moment suboptimale (238-240).

2.6. Article

YORDANOV, Y ; Dechartres, A ; Porcher, R ; Boutron, I ; Altman, DG ; Ravaud, P,
Avoidable waste of research related to inadequate methods in clinical trials., BMJ, 2015,
350, h809



Avoidable waste of research related to inadequate methods in clinical trials

Youri Yordanov,^{1, 2} Agnes Dechartres,^{1, 3, 4} Raphaël Porcher,^{1, 3, 4} Isabelle Boutron,^{1, 3, 4, 5} Douglas G Altman,⁶ Philippe Ravaut^{1, 3, 4, 5, 7}

¹Centre de Recherche Épidémiologie et Statistique, INSERM U1153, Paris, France

²Service des Urgences, Hôpital Saint-Antoine, Assistance Publique-Hôpitaux de Paris, Paris, France

³Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, Paris, France

⁴Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

⁵French Cochrane Centre, Paris, France

⁶Centre for Statistics in Medicine, University of Oxford, Oxford, UK

⁷Department of Epidemiology, Mailman School of Public Health, Columbia University New York, USA

Correspondence to: A Dechartres, Centre de Recherche Épidémiologie et Statistique, INSERM U1153, Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, 1 place du Parvis Notre Dame, 75004 Paris, France agnes.dechartres@htd.aphp.fr

Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmj.h809>)

Cite this as: *BMJ* 2015;350:h809
doi:10.1136/bmj.h809

Accepted: 20 January 2015

ABSTRACT OBJECTIVE

To assess the waste of research related to inadequate methods in trials included in Cochrane reviews and to examine to what extent this waste could be avoided. A secondary objective was to perform a simulation study to re-estimate this avoidable waste if all trials were adequately reported.

DESIGN

Methodological review and simulation study.

DATA SOURCES

Trials included in the meta-analysis of the primary outcome of Cochrane reviews published between April 2012 and March 2013.

DATA EXTRACTION AND SYNTHESIS

We collected the risk of bias assessment made by the review authors for each trial. For a random sample of 200 trials with at least one domain at high risk of bias, we re-assessed risk of bias and identified all related methodological problems. For each problem, possible adjustments were proposed that were then validated by an expert panel also evaluating their feasibility (easy or not) and cost. Avoidable waste was defined as trials with at least one domain at high risk of bias for which easy adjustments with no or minor cost could change all domains to low risk. In the simulation study, after extrapolating our re-assessment of risk of bias to all trials, we considered each domain rated as unclear risk of bias as missing data and used multiple imputations to determine whether they were at high or low risk.

RESULTS

Of 1286 trials from 205 meta-analyses, 556 (43%) had at least one domain at high risk of bias. Among the sample of 200 of these trials, 142 were confirmed as

high risk; in these, we identified 25 types of methodological problem. Adjustments were possible in 136 trials (96%). Easy adjustments with no or minor cost could be applied in 71 trials (50%), resulting in 17 trials (12%) changing to low risk for all domains. So the avoidable waste represented 12% (95% CI 7% to 18%) of trials with at least one domain at high risk. After correcting for incomplete reporting, avoidable waste due to inadequate methods was estimated at 42% (95% CI 36% to 49%).

CONCLUSIONS

An important burden of wasted research is related to inadequate methods. This waste could be partly avoided by simple and inexpensive adjustments.

Introduction

In 2009, Chalmers and Glasziou raised an important concern about the extent of research that is wasted, estimating the loss to be as much as 85% of research investment.¹ This waste concerns all types of research and occurs at all stages of the production of research evidence, from the choice of questions that are not relevant to patients and their physicians to under-reporting of trial methods and results.¹⁻⁸ Such a situation is ethically, scientifically, and economically indefensible.^{9,10} It necessitates rethinking the whole system of clinical research to increase the value of research and reduce waste, as recently outlined in a series in the *Lancet*.²⁻⁷

A large part of waste is related to inadequate methods.^{1,6} Flaws in design, conduct, and analysis can bias results of randomised controlled trials (RCTs) and the systematic reviews that include them, thus leading to potentially erroneous conclusions⁶ with serious consequences for patients. Empirical evidence found exaggerated estimates of intervention effect in trials with inadequate sequence generation or allocation concealment,¹¹⁻¹³ lack of blinding,^{11,14} or exclusion of patients from analyses.^{15,16} On the basis of this empirical evidence, the Cochrane Collaboration developed a tool for assessing risk of bias in RCTs, the Risk of Bias (RoB) tool, and recommended excluding trials at high risk of bias from the meta-analysis or presenting the meta-analysis stratified by risk of bias.¹⁷ Risk of bias is assessed from the final RCT report, when it is no longer possible to change anything. However, a part of the waste related to inadequate methods could have been avoided by modifying the design, conduct, and analysis of these trials during the planning stage of the trial.

We aimed to assess the waste of research due to inadequate methods in clinical trials and to examine to what extent that waste could be reduced. A secondary objective was to perform a simulation study to re-estimate the avoidable waste if all trials were adequately reported.

WHAT IS ALREADY KNOWN ON THIS TOPIC

In 2009, Chalmers and Glasziou made the claim that up to 85% of research could be considered as wasted

A large part of that waste is related to inadequate methods

Flaws in design, conduct, and analysis can bias results of randomised controlled trials and the systematic reviews that include them and lead to potentially erroneous conclusions with serious consequences for patients

WHAT THIS STUDY ADDS

We found that part of the waste related to inadequate methods could have been avoided by simple and inexpensive methodological adjustments

Such adjustments could decrease the risk of bias in half of trials at high risk of bias and could transform all domains at high risk to low risk in 12% trials (95% CI 7% to 18%)

In a simulation study correcting for incomplete reporting, this avoidable waste represented 42% (95% CI 36% to 49%).

Methods

We identified all trials included in recent Cochrane reviews. For a random sample of the trials with at least one domain at high risk of bias according to the review authors, we reassessed risk of bias. Then, for trials confirmed at high risk, we identified all methodological problems and proposed adjustments. An expert panel validated these adjustments and assessed their feasibility and cost. Avoidable waste was defined as trials with at least one domain at high risk of bias for which easy adjustments with no or minor cost could change all domains to low risk of bias. Finally, we performed a simulation study to re-estimate the avoidable waste related to inadequate methods if all trials were adequately reported.

Selection of trials included in recent Cochrane reviews

Identification and selection of Cochrane reviews

We searched the Cochrane Database of Systematic Reviews between 1 April 2012 and 31 March 2013 and included new systematic reviews of randomised controlled trials, with at least one meta-analysis, examining the effects of healthcare interventions. We selected reviews that had used the RoB tool for assessing risk of bias in included trials, as recommended by the *Cochrane Handbook*.¹⁷ Updates of systematic reviews were excluded because their risk of bias was frequently not entirely reassessed. We also excluded reviews of diagnostic test accuracy, prognosis, or economic evaluations.

Meta-analysis selection

For each included review, we selected the meta-analysis for the primary outcome defined by the review authors. In cases of several primary outcomes, we selected the meta-analysis with the largest number of trials. We assessed whether the outcome was objective or subjective according to the classification by Savovic et al.¹² Objective outcomes were defined as all cause mortality, other objectively assessed outcomes (for example, pregnancy, live births, and laboratory outcomes), or objectively measured but potentially influenced by clinicians' or patients' judgment (for example, hospitalisations, total dropouts, operative delivery, additional treatments administered). Clinician assessed outcomes and patient reported outcomes were considered subjective outcomes.

Trial selection

All trials included in the meta-analyses defined above were selected.

Review authors' assessment of risk of bias

For each trial, one of us (YY) collected, from the Cochrane report, the review authors' assessment of risk of bias for the following key domains of the RoB tool: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment and incomplete outcome data. We did not consider the domains "selective outcome reporting" and "other risk of bias," because these two domains are difficult to assess,^{18 19} particularly for selective outcome reporting when the protocol is not available.

Identification of a random sample of trials at high risk of bias

Among all trials with at least one domain at high risk, we randomly selected a sample of 200 trials and retrieved the full text for each of these trials. When the full text was not available (n=6) or was in a language other than English, French, or Spanish (n=19), we retrieved the English abstract and the study summary available in the "Characteristics of included studies" section of every Cochrane review.

Reassessment of risk of bias for the random sample of trials considered at high risk

For each trial of the random sample of 200 trials the review authors considered at high risk of bias, two trained reviewers (CR, YY) independently re-evaluated the risk of bias for each key domain, using the RoB tool, from the individual published trial reports. Definitions used for each domain to reassess the risk of bias were based on the *Cochrane Handbook*¹⁷ and are detailed in appendix 1. Blinding of outcome assessment and incomplete outcome data domains were assessed at the outcome level and corresponded to the primary outcome assessed in the meta-analysis. We compared our re-assessment to the initial assessment by the review authors including their support for judgment to determine any discrepancies and if review authors obtained additional information from the trial authors that could explain the discrepancies. All discrepancies were discussed and a final consensus on risk of bias was achieved with the help of a third reviewer (AD) if needed. The 95% confidence intervals of the proportion of agreement between the review authors' assessment and our re-assessment were estimated using logistic regression models fitted with the generalised estimating equation (GEE) to account for within trial correlation.²⁰

Identification of all methodological problems responsible for the high risk of bias

For each trial confirmed at high risk of bias after reassessment, two authors (AD, YY) independently identified, from the individual reports, the methodological problem(s) for each domain with high risk. They organised all methodological problems by domain and by type. For each type of methodological problem, they suggested possible adjustments from their experience and from relevant literature.^{11 15 21 22}

Feasibility and costs of methodological adjustments

We used an expert consensus approach to evaluate the proposed adjustments. During a meeting, methodological problems and possible adjustments were discussed by an expert panel including four internationally recognised methodologists (DA, IB, PR, SH) who have been involved in numerous clinical trials of various fields. Two are experts in the evaluation of non-pharmacological treatments. One of us (YY) presented, domain by domain, each type of methodological problem and proposed adjustments, with detailed examples from the trials in the sample. The expert panel was not informed

of the proportion of methodological problems in included trials. For each problem, the expert panel was asked to validate the proposed adjustments by answering the following questions: “Do you agree with the proposed adjustment?” and “Do you have another suggestion?” Then, experts were asked to estimate the feasibility and cost of the adjustments in light of their experience: “According to you and based on your experience, would you consider that the proposed adjustment is easy to implement, moderately easy, difficult, or impossible in most cases.” “According to you and based on your experience, what would be the approximate cost of this adjustment: no cost defined as $\leq 1\%$ of the total amount of the trial; minor cost, defined as $\leq 5\%$; moderate cost, defined as 5% to 15%; or major cost, if 15% or more.” These percentages were indicative. Two authors (AD, YY) facilitated the meeting to ensure that all experts first gave their opinion and then discussed together, to avoid the opinion of one leading person influencing the others.

The final feasibility and cost evaluation of every proposed adjustment was based on group consensus.

As a quality assurance measure, we also discussed feasibility and cost of proposed adjustments for blinding and practical issues with experts in the field. For pharmacological treatments, we contacted the senior pharmacist in charge of the pharmaceutical services for all pharmacological clinical trials sponsored by Assistance Publique-Hôpitaux de Paris, a network of all teaching hospitals in Paris and its suburbs. For non-pharmacological treatments, we contacted two surgeons: one orthopaedic surgeon and one obstetrician involved in clinical trials. These three experts were blinded to the experts' consensus.

Avoidable waste of research related to inadequate methods

For each trial in our random sample confirmed at high risk of bias, we assessed whether the identified methodological problems could be corrected by easy adjustments,

at no or minor additional cost as determined by the expert panel. Avoidable waste was defined as trials with at least one domain at high risk of bias for which easy adjustments with no or minor cost could change all domains to low risk of bias. As a sensitivity analysis, we also assessed the avoidable waste after applying all possible adjustments (that is, adjustments considered by the expert panel as easy, moderately easy, or difficult), whatever their cost.

Correction of incomplete reporting in the whole sample of trials and re-estimation of avoidable waste

We performed a simulation study to re-estimate avoidable waste due to inadequate methods if all trials were adequately reported. Firstly, we extrapolated our reassessment of risk of bias to the trials for which risk of bias was not reassessed (whole sample minus the 200 trials already reassessed) using the observed agreement by domain between the review authors' initial assessment and our reassessment of the 200 trials for which we had both evaluations. Then, we considered each domain rated as unclear as missing data and used multiple imputations to attribute a risk-of-bias assessment of high or low to these domains. Finally, we re-estimated the avoidable waste in the whole sample of trials after applying easy adjustments with no or minor additional cost to each domain imputed at high risk of bias. Imputation and statistical analysis are detailed in appendix 2. Statistical analyses involved use of R v3.0.2 (2013-09-25) (R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>).

Results

Selection and characteristics of trials

Figure 1 shows the study profile. Briefly, of the 1017 Cochrane citations retrieved by searching, we selected 205 Cochrane reviews including 1286 trials in the meta-analysis of the primary outcome. General characteristics of the 1286 trials are given in table 1. The most common medical fields were psychiatry (18%), obstetrics and gynaecology (10%), and oncology (8%). Trials were published between 1957 and 2012 (median year of publication 2004). In 1003 trials (78%), the outcome of interest included a non-fatal event and in 302 (23%), mortality. The outcome was considered objective in 500 trials (39%).

Risk of bias of trials

Review authors' assessment of risk of bias

Overall, 207 trials (16%) had all domains at low risk of bias, 556 (43%) at least one domain at high risk of bias, and another 523 (41%) at least one domain at unclear risk. For sequence generation, 36 trials (3%) were considered at high risk of bias and 612 (48%) at unclear risk. For allocation concealment, 64 (5%) were considered at high risk of bias and 661 (51%) at unclear risk. For blinding of patients and personnel, 379 (29%) were considered at high risk of bias and 310 (24%) at unclear risk. For blinding of outcome assessment, 249 (19%) were considered at high risk of bias and 356

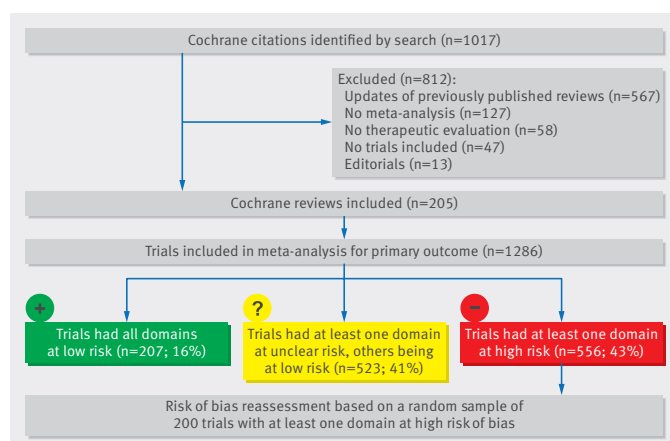


Fig 1 | Flow diagram of the selection of trials

Table 1 | General characteristics of the trials included in the meta-analysis of the primary outcome of recent Cochrane reviews

	No (%) n=1286
Medical fields:	
Psychiatry/psychology	232 (18)
Obstetrics/gynaecology	124 (10)
Oncology	100 (8)
Infectious diseases	91 (7)
Gastroenterology	73 (6)
Paediatrics	71 (6)
Others	595 (46)
Type of intervention:	
Drug	826 (64)
Counselling/lifestyle	258 (20)
Surgical/procedure	165 (13)
Equipment	37 (3)
Year of publication, median (min–max)	2004 (1957–2012)
Design of included trials:	
Parallel design	1229 (96)
Crossover	25 (2)
Cluster	21 (2)
Factorial	7 (0.5)
Non-inferiority	4 (0.3)
Primary outcome:	
Non-fatal events	1003 (78)
Physician driven data	649 (50)
Mortality	302 (23)
Patient reported outcomes	240 (19)
Biological test	186 (14)
Radiology exam	58 (5)
Subjective outcome	786 (61)
Objective outcome	500 (39)

(28%) at unclear risk. For incomplete outcome data, 220 (17%) were at high risk of bias and 243 (19%) at unclear risk (table 2).

Risk of bias reassessment for a random sample of 200 trials with at least one domain at high risk of bias

For the random sample of 200 trials with at least one domain at high risk of bias according to the review authors, we confirmed high risk of bias for 142 trials (71%). Appendix 3 presents the agreement between review authors' assessment and our re-evaluation for each domain. Disagreements with review authors' assessments mainly concerned blinding of outcome assessment. Of the 82 trials that the review authors considered at high risk for blinding of outcome assessment, 30 concerned an objective outcome so the risk of bias was judged as low. Risk of bias assessment for the 142 trials confirmed at high risk is shown in table 2.

Methodological problems identified in the trials confirmed at high risk of bias and possible adjustments

We identified 25 different types of methodological problems in the 142 trials with at least one domain confirmed at high risk. The most frequently encountered were exclusion of patients from analysis in 50 trials (35%), lack of blinding with a patient reported outcome in 27 (19%), lack of blinding when comparing a non-pharmacological intervention versus nothing in 23 (16%), and an inadequate method to deal with missing data in 22 (15%) (table 3). Table 3 describes, for each type of methodological problem, the proposed adjustments, their feasibility, and costs according to the experts.

Table 2 | Risk of bias for each key domain of the Cochrane Collaboration Risk of Bias tool. Values are number (percentage) of trials

Level of bias in key domains of the Risk of Bias tool	All trials (n=1286)	Trials with ≥1 domain at high risk of bias according to the review authors (n=556)	Trials confirmed to be at high risk of bias from a random sample of 200 trials with ≥1 domain at high risk according to the review authors (n=142)
Sequence generation (selection bias):			
Low	634 (49)	235 (42)	66 (46)
Unclear	612 (48)	285 (51)	62 (44)
High	36 (3)	36 (6)	14 (10)
Allocation concealment (selection bias):			
Low	555 (43)	177 (32)	49 (35)
Unclear	661 (51)	312 (56)	76 (54)
High	64 (5)	64 (12)	17 (12)
Blinding of participants and personnel (performance bias):			
Low	553 (43)	95 (17)	50 (35)
Unclear	310 (24)	80 (14)	10 (7)
High	379 (29)	373 (67)	82 (58)
Blinding of outcome assessment (detection bias):			
Low	631 (49)	169 (30)	64 (45)
Unclear	356 (28)	130 (23)	23 (16)
High	249 (19)	249 (45)	55 (39)
Incomplete outcome data (attrition bias):			
Low	782 (61)	240 (43)	57 (40)
Unclear	243 (19)	82 (15)	17 (12)
High	220 (17)	220 (40)	68 (48)

Table 3 | Methodological problems identified in trials confirmed to be at high risk of bias after reassessment and proposed adjustments with their feasibility and cost determined by the expert panel

Domain	Type of problem identified in the original articles	No (%) n=142	Proposition of methodological adjustments	Feasibility of adjustment (easy, medium, difficult, or impossible)	Cost (no cost, minor, moderate, or major)
Sequence generation	Inappropriate sequence generation	15 (11)	Computer sequence generation	Easy	No cost
Allocation concealment	No allocation concealment or according to patients characteristics	14 (10)	Central randomisation, internet based	Easy	Minor
	Unsealed envelopes	1 (1)	Sealed opaque envelopes	Easy	No cost
Blinding of participants and personnel:					
Pharmacological interventions	No blinding, active treatment v active treatment, both pharmacological treatments have the same physical characteristics (eg, powder, infusion) and the same route. Procedures and monitoring are similar.	2 (1)	Preparation by the pharmacist of indistinguishable containers with experimental treatment and control ± reconstitution by an independent nurse	Easy	Minor
	No blinding, pharmacological treatments v nothing. Pharmacological treatment with no noticeable side effects, shape, colour or taste.	7 (5)	Preparation by the pharmacist of indistinguishable containers with experimental treatment and placebo	Easy	Moderate
	No blinding, pharmacological treatments with different route of administration and appearance.	13 (9)	Double dummy	Medium	Major
	No blinding due to different surveillance and adaptation (eg, adaptation of dose of treatment)	3 (2)	Sham adapted surveillance and dosage in the control group	Difficult	Major
	No blinding. Treatments with noticeable elements (colour, shape, etc)	2 (1)	Blinding of the noticeable elements of the treatments	Difficult	Major
	Treatments with very frequent and clinically specific side effects	4 (3)	Impossible in most cases	Impossible in most cases	
Non-pharmacological interventions	No blinding, surgical interventions with similar scars, perioperative and postoperative care	10 (7)	Blinding of patients, no more contact with the unblinded care providers during the trial follow-up	Medium	Minor
	No blinding, surgical interventions with different scars	2 (1)	Blinding of patients with masking of scars, no more contact with the unblinded care providers during the trial follow-up	Medium	Minor
	No blinding, non-pharmacological intervention v nothing	23 (16)	Use of an attention placebo control group, with patients blinded to the hypothesis	Medium	Major
	No blinding, non-pharmacological intervention v minimal intervention	12 (8)	Blinding of patients by a simulated interventions	Difficult	Major
	No blinding, non-pharmacological intervention (surgery) v pharmacological intervention	2 (1)	Blinding of patients and personnel by a simulated intervention and a placebo	Difficult	Major
	No blinding, non-pharmacological intervention (surgery) v another non-pharmacological intervention	7 (5)	Double sham strategies	Difficult	Major
Blinding of outcome assessor	Outcome is a clinical measure performed by an unblinded physician	10 (7)	Blinded measurement by an independent clinician	Medium	Moderate
	Outcome is a non-fatal event evaluated by the investigators following patients, no blinding	11 (8)	Blinded central evaluation of events by independent clinicians (adjudication committee)	Medium	Major
	Outcome is a patient reported outcome, treatment blinding is possible	27 (19)	Participants and personnel blinding implemented	Depends on blinding	
	Outcome is a patient reported outcome, no blinding possible	0	Outcome is a patient reported outcome, no blinding possible	Impossible in most cases	
	Outcome is evaluated by imagery without blinding	4 (3)	Blinded central evaluation of imagery	Medium	Major
	Outcome is a specific biological test requiring interpretation	2 (1)	Blinded central evaluation by an independent clinician	Medium	Major
Incomplete outcome data	Exclusion of patients from the analysis	50 (35)	Intention to treat analysis	Easy	No
	Intention to treat analysis but inadequate missing data imputation	22 (15)	Intention to treat analysis with a multiple imputation method	Easy	Minor
	Important lost to follow-up rates ≥30%	16 (11)	Better patient monitoring	Difficult	Major

Appendix 4 gives some examples of methodological problems that could not be corrected.

Avoidable waste of research related to inadequate methods

From the experts' evaluation, at least one adjustment could be made in 136 trials (96%). Applying easy methodological adjustments with no or minor additional cost to trials with at least one domain at high risk of bias could have reduced the number of domains at high risk in 71 trials (50%). These adjustments could correct all trials at high risk of bias for sequence generation (n=14) and allocation concealment (n=17). Of the 82 trials at high risk for blinding of patients and personnel, seven (8%) could be corrected. Of the 55 trials at high risk of bias for blinding of outcome assessment, one (2%) could be corrected. Of the 68 trials

at high risk for incomplete outcome data, 37 (54%) could be corrected.

These adjustments resulted in all domains becoming at low risk for 17 trials (12%). So the avoidable waste related to inadequate methods represented 12% (95% CI 7% to 18%) of trials. In another 34 trials (24%; 95% CI 17% to 31%), there was no domain at high risk but at least one domain was at unclear risk. With all adjustments considered possible by the experts, regardless of difficulty or cost, avoidable waste of research represented 31% (n=44) (95% CI 23% to 39%) of trials (fig 2).

Correction of incomplete reporting in the whole sample of trials and re-estimation of avoidable waste

After imputation of domains at unclear risk of bias in the 1286 trials, 43% of trials (95% CI 39% to 47%)

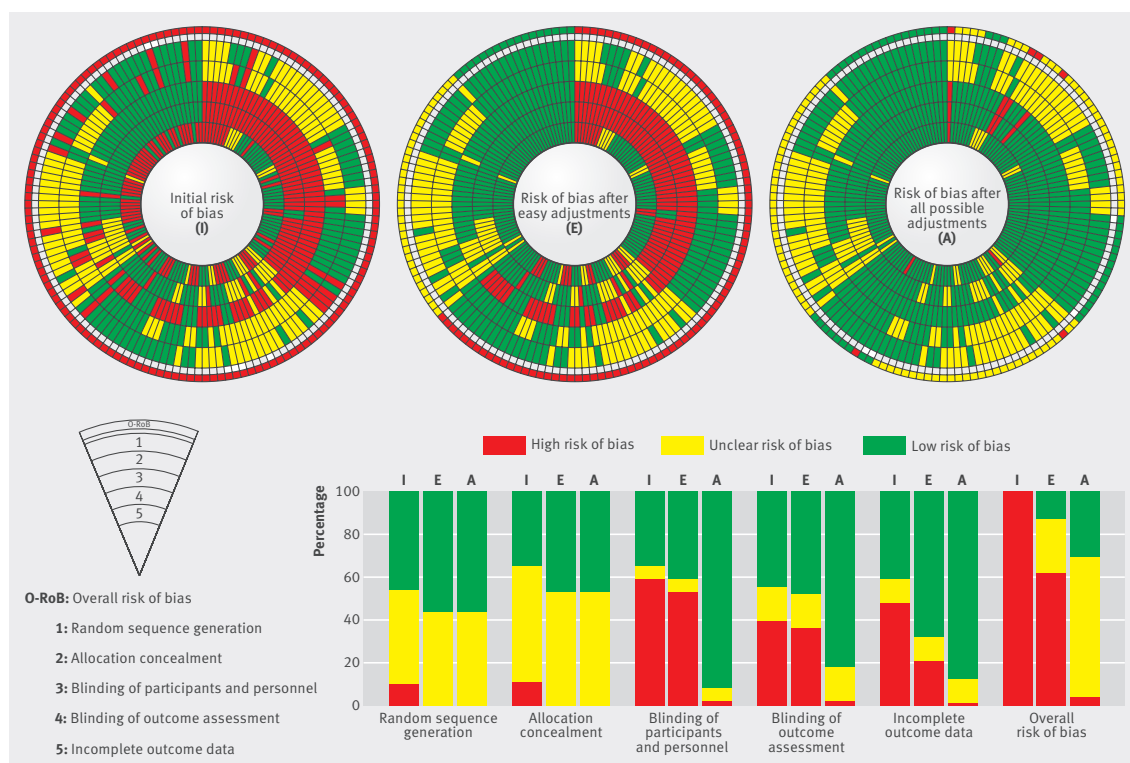


Fig 2 | Avoidable waste of research related to inadequate methods in trials with at least one domain confirmed at high risk of bias (n=142). This figure summarises the evolution in risk of bias assessment per domain for each trial, initially (I), after easy (E) methodological adjustments, or after all (A) possible methodological adjustments. Each spoke represents one of the 142 trials with at least one domain at high risk of bias. The bricks are a visual representation of the risk of bias assessment per domain, red indicating high risk of bias; yellow, unclear risk; and green, low risk for each trial. Every concentric circle represents one of the key domains of the risk of bias tool with “sequence generation” being the furthest circle from the centre, and “incomplete outcome data”, the central circle. The most external circle represents a quick overview of the risk of bias for each trial (red if at least one domain is at high risk, yellow if at least one domain is at unclear risk, and green if all domains are at low risk). The histogram is a representation of the risk of bias, per domain, across the 142 high risk trials, initially (I), after easy (E) methodological adjustments, or after all (A) possible methodological adjustments

would have all their domains at low risk and 57% (95% CI 53% to 61%) would have at least one domain at high risk. Applying easy adjustments at no or minor cost would change risk of bias to low risk for all domains in 42% of the trials with at least one domain at high risk (95% CI 36 to 49%), so the avoidable waste related to inadequate methods would represent 42% (95% CI 36 to 49%). This results in a 24% decrease of the percentage of trials with at least one domain at high risk of bias from 57% (95% CI 53% to 61%) to 33% (95% CI 29% to 37%).

Discussion

Here we assessed the avoidable waste of research related to inadequate methods in a large sample of clinical trials included in recent systematic reviews. Overall, 43% of these trials had at least one domain at high risk of bias and another 41% had at least one domain at unclear risk because of incomplete reporting. Applying easy methodological adjustments with no or minor additional cost could have limited the number of domains at high risk of bias in 50% of trials with at least one domain at high risk, with all domains changing to low risk in 12%. After correcting incomplete reporting, this avoidable waste could represent 42% of all trials with at least one domain at high risk (95% CI 36% to 49%).

Strengths and weaknesses of the study

There is a pressing need to improve the quality of research to minimise biases.⁶ Our study involved an original approach moving from an a posteriori assessment of risk of bias to limiting the methodological problems responsible for high risk of bias during the planning stage of the trials. Our approach takes into account that, for some trials, it is not possible to do better, for example, in assessing non-pharmacological treatments for which blinding is not possible. In fact, our definition of avoidable waste is based on problems that could have been easily avoided with no additional cost at the planning stage of the trial. We used an expert panel consensus to validate proposed methodological adjustments and to determine their feasibility and cost. The experts involved are internationally recognised in the field of RCT methods and assessment of risk of bias. To ensure that they were not influenced during their assessment, they were not informed of the proportion of methodological problems in the sample.

Our study focused on waste related to inadequate methods and did not consider other important sources of waste such as lack of relevance of the study question. Our sample of trials is from Cochrane reviews, which may have excluded trials not meeting certain methodological criteria. To identify methodological problems, we reassessed risk of bias for a random sample of 200 trials at high risk of bias, which may lead to a possible underestimation of the number of methodological problems. In fact, some of the trials the review authors considered at low or unclear risk could be at high risk. We disagreed with the review authors' assessment for 29% of trials. Disagreements mainly concerned the

domains related to blinding, which is consistent with previous studies showing low reproducibility for that domain.^{23 24} Although we attempted to provide a classification of all methodological problems in as much detail as possible, feasibility and costs of adjustments may vary across trials for the same methodological problem. We accounted for the correlation between the different adjustments but we did not fully take into account that for blinding; a single adjustment could simultaneously correct two domains (blinding of patients/personnel and blinding of outcome assessment). Finally, our analysis is based on the assumption that conduct and reporting are unrelated.

Possible explanations and implications for researchers

Our results highlight that with simple adjustments, the number of domains at high risk could have been reduced for half of trials with at least one domain at high risk of bias. These results may be explained by the lack of involvement of methodologists and statisticians at the planning stage of the trials and by authors' insufficient knowledge of research methods.^{6 25} In many countries, specific training in research methodology is unavailable, other than a short introduction to biostatistics early in medical school.⁶ Teaching research methods to all medical students is crucial. Not all students will be involved in research, but all will have to critically appraise research articles and make medical decisions using the results. In addition, communication between methodologists and health researchers should be enhanced by the involvement of methodologists from the planning stage of the study onwards.⁶ Access to methodologists is not always feasible for all physicians, so developing online tools allowing for diagnosis of methodological problems and proposing simple adjustments such as those we proposed would be helpful.

Our results showed that methodological adjustments led to limited improvement in terms of all domains because of the issue of incomplete reporting. Half of the trials in our study were affected by reporting issues, which confirms the poor reporting found in numerous publications.^{26–28} Although poor reporting does not mean poor methods,^{29 30} it represents another source of waste⁵ because we cannot adequately assess the validity of methods used and consequently the results and conclusions of the trial. Waste related to incomplete reporting could be completely avoided if all relevant elements were adequately noted in trial reports. In this study, when we corrected both reporting and inadequate methods, the proportion of trials with at least one domain at high or unclear risk decreased from 84% (41% with at least one domain at unclear risk and 43% with at least one domain at high risk) to 33% in the imputed dataset. This finding outlines the need to further improve reporting. Many efforts have been made to improve the reporting of research with the development and large diffusion of reporting guidelines^{31 32} adapted to the type of trial and interventions and with the creation of the Enhancing the QUALity and Transparency

Of health Research (EQUATOR) network³³ to promote transparency in research. However, according to recent publications,^{34–36} adherence to these recommendations is still suboptimal. Many journals adhere to the CONSORT statement and other reporting guidelines, but actual implementation varies greatly from one journal to another.^{35 37 38} More active implementation of the CONSORT statement and other reporting guidelines is needed to substantially improve the reporting of health related research.

Conclusions

Our study shows that an important burden of waste is related to inadequate methods in clinical trials. This waste could be avoided in part by simple and inexpensive methodological adjustments during the planning stage. Greater involvement of methodologists at the planning stage could help reduce this avoidable waste.

We thank Carolina Riveros for help with the risk of bias assessment, Sally Hopewell for participating to the expert panel, and Ali Al Khafaji for help with data extraction. We also thank Annick Tibi, pharmacist in charge of the pharmaceutical services for all clinical trials sponsored by Assistance Publique-Hôpitaux de Paris; Patrick Boyer, orthopaedic surgeon at Bichat Hospital, Paris; and Patrick Rozenberg, obstetrician at Poissy hospital for discussing feasibility and cost of methodological adjustments for blinding in pharmacological and non-pharmacological trials, respectively.

Contributors: YY was involved in the study conception, selection of trials, data extraction, data analysis, interpretation of results, and drafting the manuscript; AD was involved in the study conception, selection of trials, data extraction, data analysis, interpretation of results, and drafting the manuscript; RP was involved in the data analysis, interpretation of results, and drafting the manuscript; IB was involved in the study conception, interpretation of results, and drafting the manuscript; DA was involved in the interpretation of results and drafting the manuscript; PR was involved in the study conception, interpretation of results, and drafting the manuscript. AD is the guarantor.

Funding: This study did not receive any specific funding. The team Centre d'Epidémiologie Clinique is supported by an academic grant from the programme "Equipe espoir de la Recherche", Fondation pour la Recherche Médicale (FRM), Paris, France (No DEQ20101221475). AD is funded by the Institut National de la Santé et de la Recherche Médicale (INSERM). DGA is supported by Cancer Research UK (C5529). These funders had no role in the design and conduct of the study, collection, management, analysis, interpretation of the data, and preparation, review, or approval of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not needed.

Transparency: AD affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing: Data available upon request for academic researchers.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
- Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet* 2014;383:176–85.

- 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;383:156–65.
- Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257–66.
- Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267–76.
- Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166–75.
- Macleod MR, Michie S, Roberts I, Dirmagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101–4.
- Scott IA, Glasziou PP. Improving the effectiveness of clinical medicine: the need for better science. *Med J Aust* 2012;196:304–8.
- Liberati A. An unfinished trip through uncertainties. *BMJ* 2004;328:531.
- UK health, science, and overseas aid: not what they seem. *Lancet* 2010;376:1437.
- Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis and rheumatism* 2009;61:1633–41.
- Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38.
- Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601–5.
- Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. *JAMA* 2010;304:793–4.
- Nuesch E, Trelle S, Reichenbach S, Rutjes AWS, Bürgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 2009;339:b3244.
- Tierney JF, Stewart AL. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;34:79–87.
- Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6:e17242.
- Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66:973–81.
- Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 2006;3:e425.
- Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 2004;57:543–50.
- Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Seida JK, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.
- Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.
- Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *JAMA* 2002;287:2817–20.
- Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11:MR000030.
- Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *BMJ* 2013;347:f3755.
- Hopewell S, Dutton S, Yu LM, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723.
- Mhaskar R, Djulbegovic B, Magazini A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol* 2012;65:602–9.
- Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swannet S, et al. Bad reporting does not mean bad methods for randomised trials:

- observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22–4.
- 31 Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med* 2010;7:e1000251.
 - 32 Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol* 2011;64:718–42.
 - 33 Simera I. EQUATOR Network collates resources for good research. *BMJ* 2008;337:a2471.
 - 34 Larson EL, Cortez M. Publication guidelines need widespread adoption. *J Clin Epidemiol* 2012;65:239–46.
 - 35 Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, Zhang S, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *Journal of multidisciplinary healthcare* 2013;6:169–88.
 - 36 Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1:60.
 - 37 Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials* 2008;9:20.
 - 38 Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012;344:e4178.

© BMJ Publishing Group Ltd 2015

Appendix 1: Definitions used for assessing the risk of bias in individual randomised controlled trials

Appendix 2: Details of the simulation analysis to correct incomplete reporting in the whole sample of trials and re-estimate the avoidable waste

Appendix 3: Agreement by domain between risk of bias assessment made by review authors and our re-assessment

Appendix 4: Examples of methodological problems that could not be corrected

Appendix 1: Definitions used for assessing the risk of bias in individual randomised controlled trials.

Domain	Risk of bias	Definition
Sequence generation	Low	- Random number table - Computer random-number generator - Coin tossing - Shuffling cards or envelopes - Minimization
	High	- Sequence generated by odd or date of birth or date of admission
	Unclear	- Method used to generate sequence of randomisation not reported
Allocation concealment	Low	- Central allocation - Sequentially numbered drug containers of identical appearance - Sequentially numbered, opaque, sealed envelopes
	High	- Predictable assignment (date of birth, alternation, open random allocation schedule, unsealed envelopes)
	Unclear	- Method to maintain allocation concealment not reported
Blinding of participants and personnel	Low	- Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken - Either participants or some key personnel were not blinded but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias
	High	- No blinding or incomplete blinding and the outcome measurement is likely to be influenced by lack of blinding (i.e., subjective outcome) - Blinding of participants and personnel attempted but likely that the blinding could have been broken (differences in co-interventions among groups) - Either participants or personnel were not blinded, and the non-blinding likely to introduce bias
	Unclear	- Insufficient information to permit judgement of “low risk” or “high risk” - Insufficient information about co-interventions to assess whether lack of blinding or incomplete blinding was likely to influence the outcome
Blinding of outcome assessment	Low	- No blinding but objective outcome (i.e., mortality, biological tests) - Blinding of outcome assessor and unlikely that the blinding could have been broken
	High	- No blinding or incomplete blinding and the outcome measurement is likely to be influenced by lack of blinding (i.e., subjective outcome) - Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding
	Unclear	- Insufficient information regarding outcome assessment blinding
Incomplete outcome data	Low	- No missing outcome data - Missing data have been imputed using appropriate methods (worst-case analysis) - Missing data balanced in numbers across intervention groups with similar reasons for missing data across groups - The proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate (< 10% of the number of patients randomised or < the number of outcomes)
	High	- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups - The proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate ($\geq 10\%$ of patients randomised or $\geq$ the number of outcomes) - As-treated analysis performed with substantial departure of the intervention received from that assigned at randomisation ($\geq 10\%$ of patients randomised or $\geq$ the number of outcomes)
	Unclear	- Insufficient reporting of attrition/exclusion (i.e., number of participants randomised and analysed not stated, no reason for missing data provided)

Appendix 2: Details of the simulation analysis to correct incomplete reporting in the whole sample of trials and re-estimate the avoidable waste

We performed a simulation analysis to correct incomplete reporting in the whole sample of trials and re-assess the proportion of avoidable waste related to inadequate methods. The simulation was performed in three successive steps.

Step 1: We started by extrapolating our reassessment of risk of bias to the remaining 1 086 trials (1 286 minus the 200 trials already reassessed). We considered the domains of the non-reassessed trials as missing data and used multiple imputation by chained equations to simulate the risk of bias that would have been obtained if the trials had been reassessed. Only one imputation was performed at this step, with 10 iterations of the Gibbs sampler. The imputation model used the risk of bias of the same domain as evaluated by the review authors, the trials medical fields and the type of outcome (objective or subjective). This implies making use of the observed agreement by domain between the review authors' initial assessment and our reassessment for the 200 trials for which we had both evaluations. This simulation relied on the assumption that data were "missing at random" (MAR) (and not completely at random), that is, that the missingness was independent of what the reassessed risk of bias of a domain would have been, given the risk of bias assessed by the review authors, the medical field and the type of outcome. In our case, this was verified by design for all trials with at least 1 domain at high risk of bias, since the 200 reassessed trials were selected at random among them. For the other trials, our model assumed that the reassessment of a domain at low or unclear risk of bias would be independent of the fact that the review authors had initially rated another domain of the same trial at high risk.

Step 2: We then considered each domain rated as unclear as missing data and used multiple imputations to attribute a risk-of-bias assessment of high or low to these domains. Data were

considered MAR and were imputed by using chained equations. Variables used for multiple imputations were the risk of bias of all domains, the year of publication of the trial (by 5 years categories), the medical field and the type of outcome. Again, one imputed dataset was obtained, with 10 iterations of the Gibbs sampler. At the end of this step, we had a set of trials with no domain rated at unclear risk of bias, for only high- or low-risk trials. In our case, the MAR assumption stated that the rating of a domain of a given trial as “unclear” could be related to other characteristics of the trial and in particular to the risk of bias of other domains but did not depend on the true risk of bias (low or high) of this domain given the other variables.

Step 3: We imputed the risk of bias for each domain obtained after applying methodological adjustments for the domains that were missing because the initial rating was unclear or because the trial was not part of the sample of trials that were reassessed and corrected. Both situations were imputed simultaneously using chained equations. Again, one imputed dataset was obtained, with 10 iterations of the Gibbs sampler. Variables used for imputation were the risk of bias of each domain before and after methodological adjustments, the medical field, the type of outcome and the year of publication. As for the preceding steps, we assumed the missing data mechanism would be MAR and that the imputation model would be correctly specified. In our case this meant that the probability to implement methodological adjustments leading to a low risk of bias specific domain did not depend on our ability to implement such adjustments; given the risk of bias of this and other domains and our ability to implement methodological adjustments to these other domains.

Steps 2 and 3 took advantage of the potential multiple correlations among domains (eg, random sequence generation and allocation concealment could always be corrected easily and at no cost, and modifying the risk of bias for both blinding domains appeared to be correlated).

The three steps were repeated 999 times, thus yielding 999 imputed datasets with no domain rated as unclear risk (all domains being at high or low risk of bias). Among these datasets, domains at high risk of bias were potentially corrected to low risk because of easy adjustments at no or minor extra cost.

The proportion of trials with at least 1 domain at high risk of bias that could change to low risk was computed for each dataset, and a pooled estimate was obtained by Rubin's rule¹. Multiple imputations involved the "mice" package² for R software³.

References

1. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York, 1987.
2. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Soft*. 2011;45(3):1-67.
3. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL <http://www.R-project.org/>.

Appendix 3: Agreement by domain between the risk of bias assessment made by the review authors and our re-assessment.

Re-assessed risk of bias, by domain n (%*)			
Initial risk of bias, by domain			
Random sequence generation (selection bias)	Low (n=93)	Unclear (n=93)	High (n=14)
Low (n=82)	79 (96)	3 (4)	0 (0)
Unclear (n=102)	13 (13)	89 (87)	0 (0)
High (n=16)	1 (6)	1 (6)	14 (88)
Agreement	0.91 (0.88-0.95)		
Allocation concealment (selection bias)	Low (n=73)	Unclear (n=114)	High (n=16)
Low (n=65)	64 (98)	1 (2)	0 (0)
Not evaluated (n=1)	0 (0)	1 (100)	0 (0)
Unclear (n=112)	4 (4)	108 (96)	0 (0)
High (n=22)	2 (9)	4 (18)	16 (73)
Agreement	0.95 (0.93-0.98)		
Blinding of participants and personnel (performance bias)	Low (n=69)	Unclear (n=49)	High (n=82)
Low (n=43)	43 (100)	0 (0)	0 (0)
Not evaluated (n=3)	1 (33)	1 (33)	1 (33)
Unclear (n=25)	10 (40)	14 (56)	1 (4)
High (n=129)	15 (12)	34 (26)	80 (62)
Agreement	0.69 (0.61-0.74)		
Blinding of outcome assesement (detection bias)	Low (n=112)	Unclear (n=33)	High (n=55)
Low (n=70)	66 (94)	0 (0)	4 (6)
Not evaluated (n=3)	1 (33)	1 (33)	1 (33)
Unclear (n=45)	15 (33)	25 (56)	5 (11)
High (n=82)	30 (37)	7 (9)	45 (55)
Agreement	0.69 (0.63-0.75)		
Incomplete outcome data (attrition bias)	Low (n=96)	Unclear (n=36)	High (n=68)
Low (n=83)	75 (90)	2 (2)	6 (7)
Not evaluated (n=7)	2 (29)	5 (71)	0 (0)
Unclear (n=37)	8 (22)	21 (57)	8 (22)
High (n=73)	11 (15)	8 (11)	54 (74)
Agreement	0.78 (0.72-0.84)		
Trial Level			
At least 1 domain at risk	Low (n=7)	Unclear (n=51)	High (n=142)
High (n=200)	7 (4)	51 (26)	142 (71)
Agreement	0.71 (0.64-0.77)		

* Percentages are calculated by row

Appendix 4: Examples of methodological problems that could not be corrected.

Study Id	Intervention	Contol	Assessed outcome	Blinding of patients/ care providers	Methodological issue related to blinding of patients/ care providers
Bujko 2006	Preoperative irradiation for rectal cancer	Radiochemotherapy, with two courses of 5-fluorouracil and leucovorin prior surgery	Local recurrence	No blinding	No blinding, treatments with very frequent and clinically specific side effects: when some people receive chemotherapeutic agents and some do not we considered that this can be a potential source of bias, and due to the very specific and frequent side effect of chemotherapy this problem was thus impossible to solve with a methodological adjustment
Gerard 2006	Preoperative radiotherapy in resectable T3-4, Nx, M0 rectal adenocarcinoma	Preoperative radiotherapy , and chemotherapy with fluorouracil and leucovorin, in resectable T3-4, Nx, M0 rectal adenocarcinoma	Local recurrence	No blinding	No blinding, treatments with very frequent and clinically specific side effects: when some people receive chemotherapeutic agents and some do not we considered that this can be a potential source of bias, and due to the very specific and frequent side effect of chemotherapy this problem was thus impossible to solve with a methodological adjustment
Blackburn 1981	Interruption of tranquilizing medication among schizophrenics patients	Patients usual treatment was replaced by placebo tablets, as placebo was physically different from the regular treatment, patients were told they were under a new drug.	Relapse, at anytime	Placebo was used (named CAT)	Double, thiamine chloride was used as placebo, participants and nurses were told that a new medication was given, but "It was noted, however, that nursing personnel were soon referring to CAT as a placebo"
Morton 1968	Chlorpromazine or trifluoperazine withdrawal in chronic schizophrenia	Patients' usual treatment was replaced by placebo tablets. Placebo was identical tablets, with no taper period	Relapse, at anytime	Placebo was used. Identical tablets, prepared by the hospital pharmacist, were used	Trial authors reported that nurses made correct guesses as to who was on drug and who was on placebo.

3. Gaspillage, évitable, de la recherche lié aux critères de jugement dans les essais cliniques

Après s'être intéressé au gaspillage, évitable, de la recherche dû à l'emploi de méthodes inadéquates dans les essais cliniques, dans cette seconde partie de la thèse nous avons cherché à évaluer le gâchis lié au choix et à la présentation des critères de jugement dans les ECRs. Ce travail a donné lieu à deux articles.

3.1 Justification de l'étude

Afin d'aider les patients et les soignants, les essais contrôlés randomisés doivent évaluer les critères de jugement les plus importants pour eux (29, 67, 141, 146, 241-244). Les chercheurs sont donc fortement encouragés à choisir, à la phase de planification, des critères de jugement importants pour les patients (143-148). Malheureusement ils échouent souvent à le faire et leur utilisation semble perfectible dans les essais cliniques de nombreux domaines (149-151, 166).

Hormis le choix des critères de jugement, la présentation sélective et/ou incomplète est une autre menace qui pèse sur leur utilité pour éclairer les patients et les professionnels de santé. Plus d'un tiers d'entre eux seraient présentés de manière sélective, sur la base de la nature et de la direction des résultats observés (57, 59, 192-195). Et 30 à 50 % seraient rapportés de manière incomplète, rendant leur inclusion dans les méta-analyses impossibles et conduisant ainsi à une vision tronquée des éléments de preuves disponibles (57, 197, 198).

Ces problèmes affectant les critères de jugement ont souvent été décrits (57, 59, 143-148, 192-195). Mais aucune étude n'a évalué la part évitable de ce gaspillage de la recherche si les chercheurs avaient mesuré, et rapporté de manière transparente, l'ensemble des critères de jugement importants.

3.2. Objectifs

Le but de cette étude était d'évaluer l'importance des critères de jugement des revues Cochrane puis de décrire le gaspillage de la recherche liée à une planification inadéquate, une présentation sélective ou incomplète des critères de jugement dans les ECRs inclus dans ces revues et enfin d'examiner dans quelle mesure ce gaspillage pourrait être évité.

3.3. Méthodes

Nous avons mené, en quatre étapes, une revue méthodologique d'ECR inclus dans des revues systématiques Cochrane.

1) Premièrement, nous avons cherché à évaluer si les critères de jugement présentés dans le tableau restituant les principaux résultats des revues Cochrane (« Summary of Findings tables » ou SoF table) étaient des critères de jugement importants pour les patients (CJIP). 2) Deuxièmement, pour tous les critères de jugement des SoF table, nous avons identifié les ECRs qui ont été inclus dans les revues, mais exclus des méta-analyses correspondantes parce que le critère de jugement manquait. 3) Ensuite, et afin d'identifier les causes de ces exclusions, nous avons systématiquement recherché les fiches d'enregistrement et les protocoles de ces ECR

pour distinguer les critères de jugement non planifiés (planification inadéquate : non inclus dans les registres ou les protocoles) de ceux qui l'étaient. Pour ces derniers, nous avons établi une distinction entre ceux qui étaient correctement rapportés, et ceux qui ont été non rapportés (présentation sélective) ou rapportés de manière incomplète (présentation incomplète). 4) Enfin, nous avons évalué, par consensus d'experts, la faisabilité et le coût de la mesure des critères de jugement non prévus (planification inadéquate).

Source des données

Nous avons obtenu de la collaboration Cochrane l'ensemble des données incluses dans les revues publiées sur période de plus de deux ans, sous forme de fichiers XML issus du logiciel RevMan (245). Nous avons sélectionné de manière aléatoire 300 revues parmi les revues avec un SoF, qui incluaient au moins un ECR publié après 2007 et qui évaluaient une intervention thérapeutique. Après avoir identifié le SoF de la comparaison principale pour chaque revue, nous avons extrait les caractéristiques générales des revues, de la comparaison principale (intervention, comparateur, patients, contexte) et les critères de jugement qui y étaient rapportés.

Nous avons émis a priori l'hypothèse que les critères de jugement présentés dans le « Summary of Findings tables » étaient les critères de jugement les plus importants (CJI) à mesurer pour une intervention thérapeutique donnée et dans un contexte particulier (29, 246). En effet, selon le Cochrane Handbook, les auteurs de revues doivent présenter sous forme de tableau les principaux résultats et les critères de jugement les plus importants de la revue (246). Ces critères

de jugement doivent être choisis au cours de la phase de planification de la revue et non en fonction de leur disponibilité ou pas dans les essais individuels.

Classification des critères de jugement

Nous avons catégorisé chaque critère de jugement du SoF selon la classification suivante : mortalité, autres événements cliniques (e.g. infarctus du myocarde), décision thérapeutique (e.g. transfusion), fonction (e.g. anxiété, incapacité et dyspnée), douleur, qualité de vie, effets indésirables, paramètres physiologiques, paramètres biologiques, paramètres radiologiques, procédure (e.g. durée d'une intervention chirurgicale), utilisation des ressources (e.g. hospitalisation), rapport coût-efficacité et satisfaction des patients. La classification a été faite en double et de manière indépendante pour 10 % du panel.

Qualité des preuves disponible

Nous avons extrait des SoF l'évaluation de la qualité des éléments de preuves disponibles, pour chaque critère de jugement, selon GRADE (Grading of Recommendations Assessment, Development and Evaluation) (18, 29, 246-249).

1) Classification des critères de jugement des SoF tables

Nous avons défini les critères de jugement importants pour les patients (CJIP) comme étant : la mortalité, les autres événements cliniques, les événements indésirables, la fonction, la douleur, la qualité de vie ou encore les décisions thérapeutiques (149, 250-252).

2) Identification des ECRs exclus des méta-analyses

Au sein de chaque revue Cochrane nous avons d'abord identifié l'ensemble des ECRs inclus dans la comparaison principale (c.-à-d. les essais inclus dans au moins une méta-analyse d'un des critères de jugement de la SoF table). Puis, pour chacun de ces essais, nous avons manuellement recherché s'ils contribuaient à toutes les méta-analyses des CJI.

3) Raisons de l'absence des critères de jugement importants des essais cliniques

Nous avons a priori émis l'hypothèse que l'exclusion d'un ECR d'une méta-analyse pourrait être liée à une planification inadéquate (c.-à-d. le CJI n'était pas prévu), à une présentation sélective (c.-à-d. le CJI était prévu, mais non rapporté dans la publication) ou une présentation incomplète (c.-à-d. le CJI a été rapporté dans la publication, mais pas de manière à permettre l'inclusion de l'essai dans la méta-analyse) (167).

Pour chaque ECR publié après 2010, et afin de pouvoir différencier les deux dernières situations, nous avons recherché les protocoles, fiches d'enregistrement, et toutes autres informations pertinentes. Pour ce faire, nous avons analysé tous les articles, résumés de conférences, rapports, etc. liés à l'essai identifié. Nous avons examiné tous les articles à la recherche d'informations relatives à l'enregistrement (nom du registre et/ou numéro d'enregistrement) et/ou à la disponibilité du protocole. Lorsqu'aucune référence à un registre n'était retrouvée, nous avons utilisé l'approche suivante :

1. Nous avons recherché dans le texte intégral de la revue Cochrane toute information concernant l'enregistrement de l'essai.
2. Puis nous avons consulté la plateforme du registre international des essais cliniques (ICTRP) de l'Organisation mondiale de la santé (OMS). De plus, selon l'affiliation de l'auteur, nous avons effectué une recherche dans le registre local s'il ne faisait pas partie du programme de l'OMS. Nous avons utilisé le titre de l'article, les noms des premiers et derniers auteurs, ou l'affiliation comme mots-clés.
3. Si aucune information n'était trouvée ainsi, nous avons cherché dans Google le titre de la publication et les mots-clés relatifs à l'enregistrement (par exemple, registre, enregistrement, NCT, etc.).
4. Enfin, si aucune information n'était trouvée, nous avons contacté l'auteur correspondant.

Puis, nous avons classé, chaque CJI, de chaque ECR exclus d'une méta-analyse, en 5 catégories :

- Planification inadéquate : critère de jugement non prévu selon les documents retrouvés ;
- Présentation incomplète : critère de jugement prévu, mais pas de manière à permettre l'inclusion de l'ECR dans la méta-analyse ou critère de jugement non prévu, mais rapporté dans le rapport, mais pas de manière à permettre son inclusion ;
- Présentation sélective : critère de jugement prévu, mais non rapporté ;
- Distinction impossible entre présentation incomplète / sélective et planification inadéquate : pas de registre ou autres documents retrouvés ;
- Autres situations.

La classification a été faite en double et de manière indépendante pour 10 % du panel.

4. Évaluation du gaspillage évitable de la recherche

À l'aide d'un consensus d'expert, nous avons évalué la difficulté et le coût éventuel de mesurer les CJI manquants. Les experts, 5 méthodologiques et investigateurs d'essais cliniques (AD, IB, PC, PR, VT), ont évalué pour chaque essai les CJI classés précédemment comme non prévus (planification inadéquate). Ils ont eu à estimer si les CJI manquants auraient pu être mesurés de manière « simple, modérément simple, difficile ou impossible dans la majorité des cas », de la perspective du chercheur ou des patients et à quel coût (aucun coût, coût mineur, coût modéré

ou coût majeur). Enfin, ils ont eu à évaluer sur le CJI était effectivement un critère de jugement important ou pas.

Nous avons défini le gaspillage évitable de la recherche, lié aux critères de jugement, comme les situations relevant de la présentation incomplète ou sélective des résultats, ou la planification inadéquate si le critère de jugement aurait pu être mesuré simplement et avec un coût mineur (aucun coût ou coût mineur). Au niveau des ECRs, le gaspillage de la recherche aurait pu être partiellement évité si l'essai avait été inclus dans au moins une méta-analyse supplémentaire. Et le gaspillage aurait pu être totalement évité si l'ECR avait pu être inclus dans toutes les méta-analyses de la comparaison principale de la revue systématique. Les analyses statistiques ont été réalisées en utilisant le logiciel R V 3.0.2. (<http://www.r-project.org>).

3. 4. Résultats

Entre mars 2011 et septembre 2014, 2796 revues systématiques Cochrane ont été publiées, 1670 n'incluaient pas de SoF (60 % des revues examinées) et 820 étaient éligibles à l'inclusion. Nous avons identifié un échantillon aléatoire de 300 revues, mais 10 ont été exclues parce que leur SoF présentait différentes interventions, ce qui a laissé 290 revues pour évaluation. Les revues comprenaient une médiane de 11 ECR. Au sein de ces revues, les comparaisons principales incluaient une médiane de 5 essais. L'intervention expérimentale était non pharmacologique dans 40 % des revues (n = 115). Chaque SoF comprenait une médiane de 5 critères de jugement, pour un total de 1414. Les méta-analyses des critères de jugement comprenaient une médiane de trois essais (Q1-Q3 : 2-7). Parmi les 1414 critères de jugement rapportés, les plus courants

étaient les critères de jugement fonctionnels (27 %, n = 384), les événements cliniques (14 %, n = 198) et les événements indésirables (12 %, n = 174). La mortalité représentait 10 % (n = 138) des résultats et la qualité de vie 7 % (n = 98). Les paramètres biologiques représentaient 6 % (n = 89), l'utilisation de ressources ou les procédures 5 % (n = 74) et les paramètres physiologiques 4 % (n = 56) des critères de jugement. Le processus de sélection des revues systématiques, des essais cliniques et enfin des critères de jugement est décrit dans la figure 12.

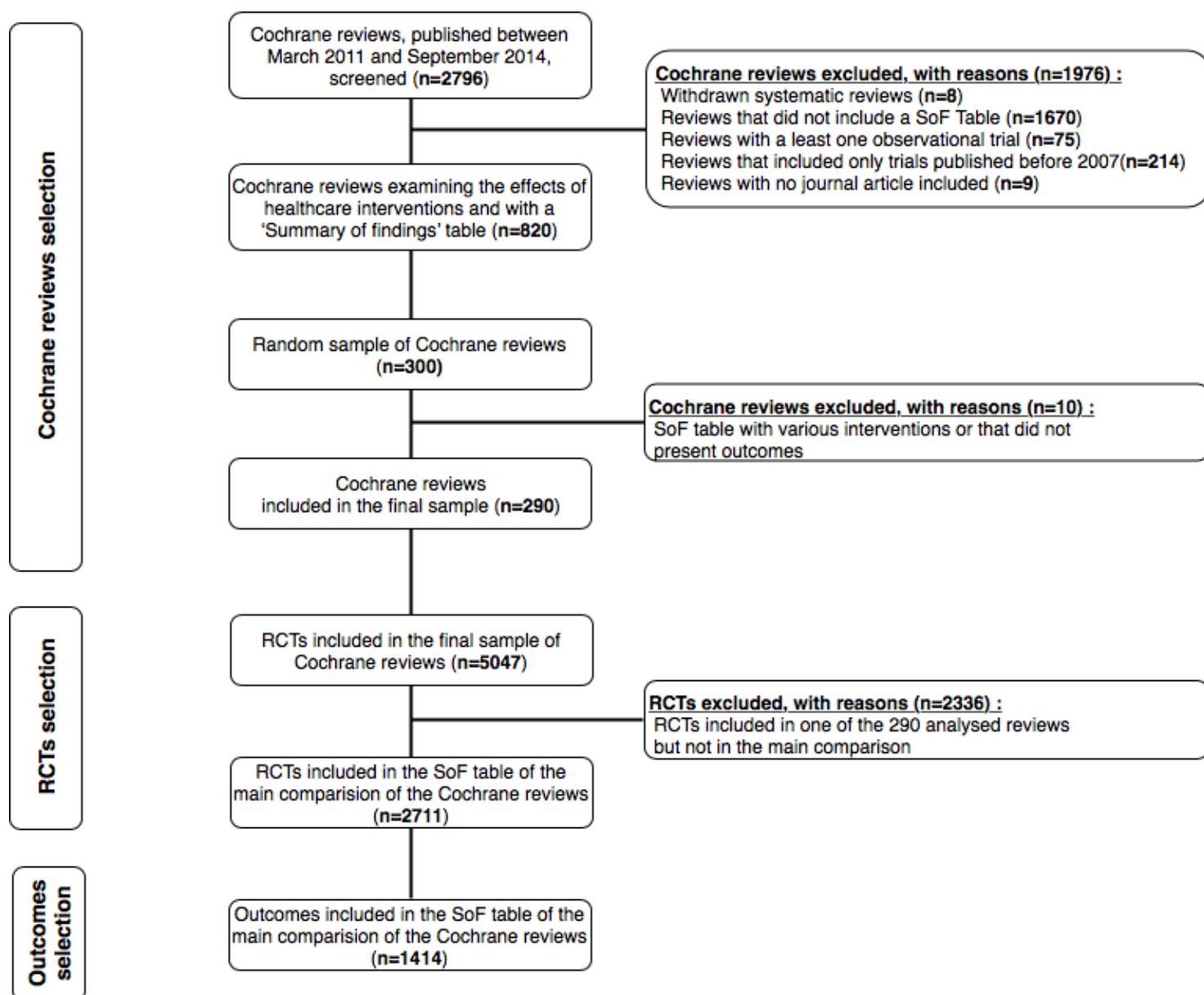


Figure 12. Organigramme du processus de sélection.

1. Classification des critères de jugement des SoF tables (Résultats extraits de Yordanov *et al*, PLoS One, 2018(253))

Un total de 1089 critères de jugement (77 %, IC 95 % : 75-79) pouvait être considéré comme des CJIP. Presque toutes les revues incluaient au moins un critère de jugement important pour les patients dans le SoF (n = 278, 96 %, IC 95 % : 93-98) (fig. 13).

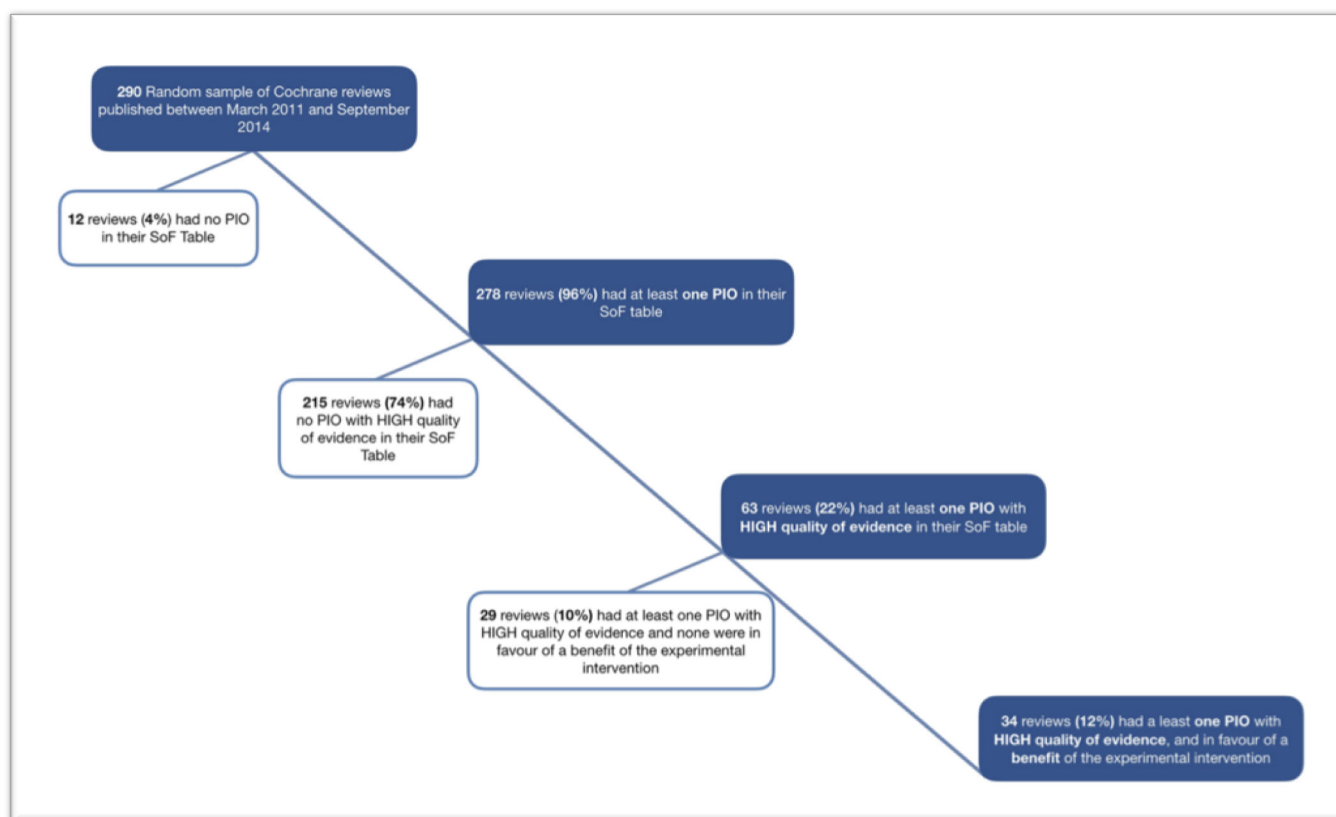


Figure 13. Critères de jugement important pour les patients dans des revues Cochrane récentes

Mais la qualité des éléments de preuve disponibles était élevée pour seulement 12 % (n=168) des critères de jugement et modérée pour 28 % (n=402). Dans 45 % des cas, celle-ci étaient

faibles à très faibles (n=640). Les résultats étaient similaires quand on se concentrait sur les CJIP, avec 12 % de ceux-ci avec une qualité des preuves disponibles estimée comme élevée. Moins du quart des revues (22 %, n = 63), comprenait au moins un CJIP avec une qualité élevée des preuves. Dans la moitié (n = 34) de ces revues, les résultats étaient en faveur de l'intervention expérimentale (fig. 13).

2. Identification des ECRs exclus des méta-analyses (Résultats extraits de Yordanov *et al*, BMC Med, 2018 (254))

Les 290 revues ont inclus au total 2711 ECR dans leur comparaison principale ; 596 (22 % ; IC 95 % : 20-24) ont été inclus dans toutes les méta-analyses des CJI et 2115 (78 %, IC 95 % : 76-80) ont été exclus d'au moins une méta-analyse.

Chaque ECR a contribué, en moyenne, à 55 % des méta-analyses pour les CJI.

3. Raisons de l'absence des critères de jugement importants des essais cliniques (Résultats extraits de Yordanov *et al*, BMC Med, 2018 (254))

Parmi les 2115 ECR exclus d'au moins une méta-analyse des CJI, 310 ont été publiés en 2010 ou plus tard. Nous avons exclu 19 ECR (articles rétractés, ou dans des langues autres que l'anglais et le français), et ainsi évalué 291 ECR, pour un total de 971 CJI manquants. Pour près de la moitié, nous avons retrouvé un numéro d'enregistrement ou un protocole (54 %, n = 156/291), correspondant à 461 CJI manquants. Les raisons de l'absence des critères de jugement importants des essais cliniques étaient une présentation incomplète pour 21 % des CJI

manquants (n = 204/971) dans 40 % des ECRs (n = 117/291) et une planification inadéquate pour 29 % des CJI manquants (n = 282/971) dans 42 % d'ECR (n = 123/291) (tableau 3.). La présentation sélective représentait 4 % des CJI manquants (n = 36/971) dans 9 % des ECRs (n = 25/291). Néanmoins, 42 % des ECRs (n = 122/291) n'avaient pas de protocole ou d'enregistrement, donc pour ceux-ci, nous ne pouvions pas faire la distinction entre une présentation sélective et une planification inadéquate. En limitant notre description aux ECR enregistrés (n = 156), une planification inadéquate concernait 61 % des CJI manquants (n = 282/461) et affectaient 79 % des ECRs (n = 123/156), les rapports incomplets représentaient 21 % des critères de jugement (n = 98/461) dans 41 % des ECRs (n = 64/156) et la présentation sélective, 7 % des résultats manquants (n = 34/461) dans 15 % des ECRs (n = 23/156) (tableau 3.).

Reasons for missing outcome	All trials published in 2010 or later			
	All trials (registered and unregistered)		Registered trials	
	No. of affected outcomes (%) N = 971	No. of affected trials (%) [*] N = 291	No. of affected outcomes (%) N = 461	No. of affected trials (%) [*] N = 156
Inadequate planning	282 (29)	123 (42)	282 (61)	123 (79)
Selective reporting	36 (4)	25 (9)	34 (7)	23 (15)
Incomplete reporting	204 (21)	117 (40)	98 (21)	64 (41)
Unable to distinguish between selective reporting and lack of planning	363 (39)	122 (42)		
Other situations	86 (9)	63 (24)	47 (10)	41 (26)

^{*}The total exceeds 100% because some outcomes were included in more than one category

Tableau 3. Raisons de l'absence des critères de jugement importants des essais cliniques

4. Évaluation du gaspillage évitable de la recherche (Résultats extraits de Yordanov *et al*, BMC Med, 2018 (254))

Nous avons soumis les 282 CJI manquants à cause d'une planification inadéquate, issus de 123 ECR, à notre panel d'experts. L'importance de ces CJI a été confirmée par nos experts dans 78 % des cas (n = 221/282). Et pour 82 % (n = 182) ils auraient pu être simplement mesurés et pour un coût mineur (appendice 4, tiré-à-part du présent article (254)). Donc, pour les 291 ECR évalués, et en tenant compte de la présentation incomplète ou sélective des CJI, le gaspillage de la recherche aurait pu être partiellement évité dans 43 % (n = 126) des essais et totalement pour 12 % (n = 34) (fig. 13). Si nous considérons également les CJI qui auraient pu être simplement mesurés et pour un coût mineur, le gaspillage de la recherche aurait pu être partiellement évité pour 63 % des ECRs (n = 183) et totalement pour 30 % (n = 86) (fig. 13).

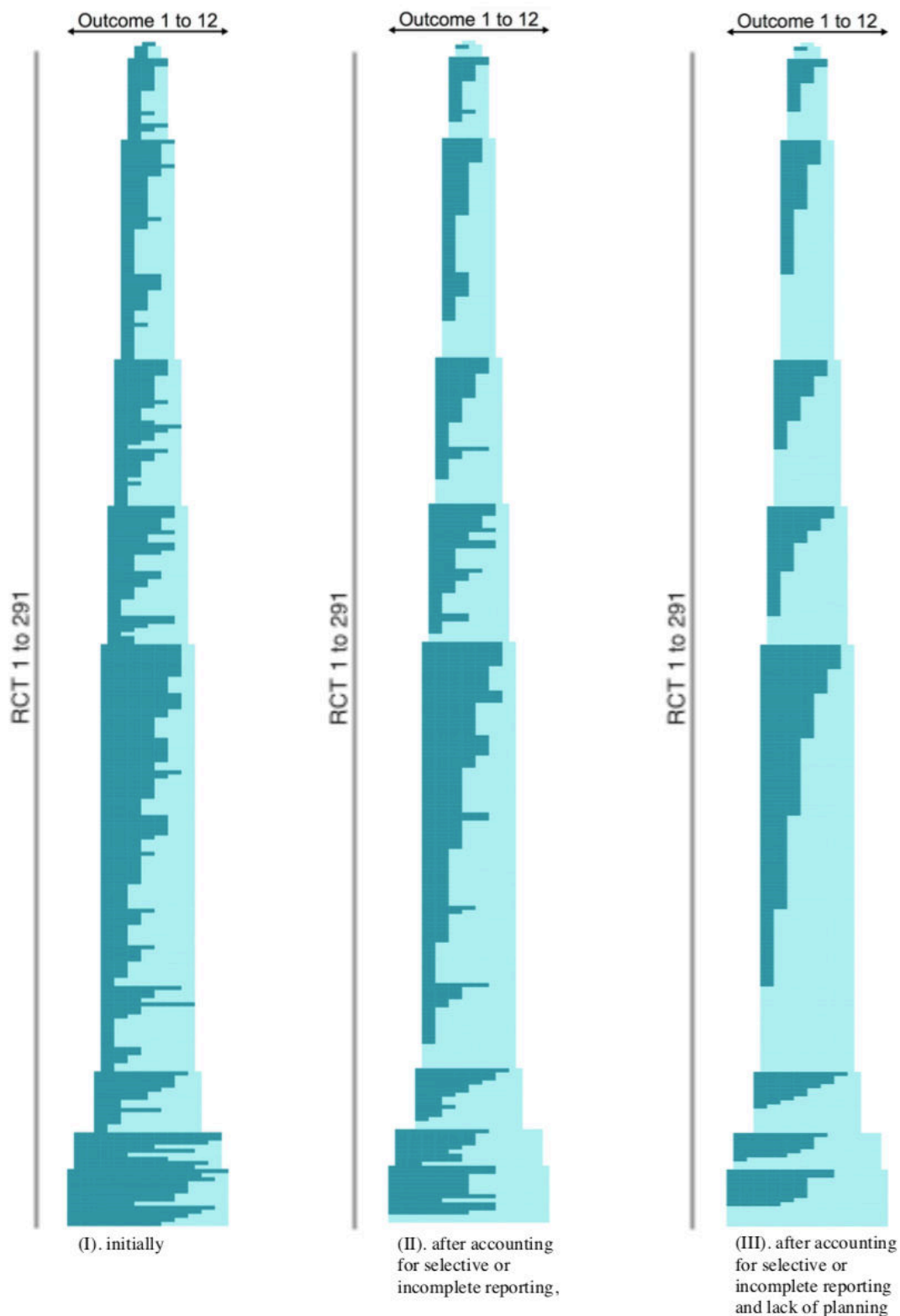


Figure 14. Gaspillage évitable de la recherche lié aux critères de jugement importants, manquants, dans les ECRs publiés en 2010 ou après (n = 291). Cette figure résume l'évolution de l'état (présent/absent) des critères de jugement des SoF table, pour chaque ECR, aux différentes étapes de notre étude. Chaque brique représente un critère de jugement. La couleur est une représentation visuelle de la présence ou de l'absence de celui-ci dans les ECRs. Le bleu clair signifie que le résultat était présent et bleu foncé que le résultat était absent dans l'ECR. Sur l'axe des Y, chaque ligne représente l'un des 291 ECR publiés en 2010 ou plus tard, et ne contribuant pas à toutes les méta-analyses des critères de jugement rapportés dans le SoF table. L'axe X représente les critères de jugement des SoF table.

3.5. Discussion

Dans cette étude, nous avons évalué les critères de jugement de revues Cochrane d'un grand nombre de spécialités médicales différentes. Plus des trois quarts des critères de jugement rapportés dans les SoF pouvaient être considérés comme des critères de jugement importants pour les patients, avec une médiane de 86 % des critères de jugement important par « Summary of Findings table ». Nos résultats suggèrent également que de nombreux essais (78 %) n'ont pas été inclus dans l'ensemble des méta-analyses des critères de jugement importants, et que chaque ECR n'a contribué qu'à la moitié des méta-analyses, en moyenne. Parmi les ECRs enregistrés, la planification inadéquate était la raison la plus souvent retrouvée (79 % des cas). Ce gaspillage aurait pu être partiellement évité pour 63 % des essais et totalement pour 30 % d'entre eux.

Nos résultats semblent conforter notre hypothèse de départ, avec une forte proportion de critères de jugement importants au sein des « Summary of Findings tables ». Ils sont proches de ceux retrouvés dans la littérature(250-252, 255). Par exemple, Ameer *et al*, rapportent que 95 % des revues Cochrane incluaient au moins un CJIP parmi leurs critères de jugement principaux, cette proportion étant plus élevée que dans les revues systématiques non Cochrane(250). Mais, dans une revue méthodologique dans le champ de la néonatalogie, Ioannidis *et al* ont analysé un grand nombre d'essais et de revues systématiques afin d'évaluer la présence d'un critère de jugement considéré comme d'une importance majeure par des experts de la spécialité. Ils rapportent que celui-ci était retrouvé dans 75 % des revues, mais dans uniquement 20 % des ECRs (255). Ces résultats pointent, encore une fois, la différence entre les critères de jugement des revues systématiques et des essais individuels, ce qui pourrait participer à expliquer la faible

qualité des preuves disponibles selon nos résultats (dans notre étude, seulement 22 % des revues avaient au moins un CJIP avec un niveau de preuve élevé).

Ensuite, l'identification des critères de jugement importants, ou les plus importants, à mesurer dans une situation est difficile et sans un consensus sur ceux-ci cette hétérogénéité persistera (256). Divers projets, ayant pour but d'améliorer la pertinence, la cohérence, l'analyse et la présentation des critères de jugement en recherche clinique, ont émergé, comme l'initiative COMET (Core Outcome Measures in Effectiveness Trials) par exemple (147, 148, 242, 257, 258). Elle propose le développement d'un ensemble minimal de critères de jugement qui devrait être mesuré et rapporté dans tous les essais cliniques portant sur une pathologie donnée et dans un contexte clinique particulier. Une autre possibilité serait d'utiliser les critères des SoF tables des revues Cochrane, la plupart semblant importants pour les patients (253). Une autre raison fréquemment retrouvée était la présentation incomplète des critères de jugement. Il est aisé d'imaginer que cette cause serait facilement corrigeable car uniquement liée à un problème d'écriture du rapport d'étude. Par exemple, les résultats des ECRs sont plus précisément rapportés sur ClinicalTrials.gov que dans les publications, probablement en raison du modèle normalisé utilisé pour rapporter les résultats sur le registre en ligne (259). Enfin, la présentation sélective des résultats semblait affecter moins d'essais dans notre panel que dans la littérature, qui décrit des taux de divergences entre les critères de jugement principaux et les protocoles ou registres de près de 30 % (57, 193). Cette différence pourrait s'expliquer par le fait que nous ayons considéré tous les critères de jugement, sans différencier les critères primaires et secondaires.

Notre étude décrit que les critères de jugement rapportés dans les tableaux SoF de récentes revues Cochrane peuvent être considérés comme des critères de jugement importants pour les patients. Mais la plupart des essais inclus dans notre échantillon n'ont pu contribuer à toutes les méta-analyses des critères de jugement importants, principalement en raison d'une planification inadéquate ou de rapports incomplets. Ce gaspillage aurait pu être partiellement évité pour 63 % des essais et totalement pour 30 %. Nous devons accélérer le développement et la diffusion des COS, et réduire la mauvaise présentation des critères de jugement pour éviter ce gaspillage.

3.1.6. Articles

YORDANOV, Y ; Dechartres, A ; Ravaud, P, *Patient-important outcomes in systematic reviews: Poor quality of evidence.*, PLoS One, 2018, 13, e0195460

YORDANOV, Y ; Dechartres, A ; Atal, I ; Tran, VT ; Boutron, I ; Crequit, P & al, *Avoidable waste of research related to outcome planning and reporting in clinical trials.*, BMC Med, 2018, 16,

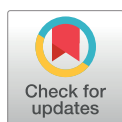
RESEARCH ARTICLE

Patient-important outcomes in systematic reviews: Poor quality of evidence

Youri Yordanov^{1,2,3,4*}, Agnes Dechartres^{1,4,5,6}, Philippe Ravaud^{1,4,5,6,7}

1 INSERM, U1153, Paris, France, **2** Sorbonne Universités, UPMC Paris Univ-06, Paris, France, **3** Service des Urgences - Hôpital Saint Antoine, Assistance Publique-Hôpitaux de Paris (APHP), Paris, France, **4** Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique-Hôpitaux de Paris (APHP), Paris, France, **5** Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, **6** Cochrane France, Paris, France, **7** Columbia University, Mailman School of Public Health, Department of Epidemiology, New York, United States of America

* youri.yordanov@aphp.fr



Abstract

Background

Cochrane reviewers are strongly encouraged to evaluate the quality of evidence for the most important outcomes by using the GRADE approach and to report these results in a Summary of Findings (SoF) table. We aimed to assess whether outcomes reported in the SoF table of Cochrane reviews could be considered patient-important outcomes (PIOs) and the quality of the available evidence for these outcomes.

Methods

We performed a methodological review of Cochrane reviews published between March 2011 and September 2014. For a random sample of Cochrane reviews reporting a SoF table, we extracted all outcomes reported in this table and evaluated whether they could be considered PIOs (i.e., mortality, other clinical events, adverse events, function, pain, quality of life and therapeutic decisions). Then, we collected the quality of evidence for every outcome in these SoF tables.

Results

We included 290 reviews issued by 47 of the 53 Cochrane Review Groups. Every SoF table included a median of 5 outcomes, for a total of 1414 outcomes; 1089 (77%) could be considered PIOs. Almost all reviews ($n = 278$, 96%) included at least one PIO in their SoF table. The quality of evidence for the outcomes was high for 12% ($n = 168$), moderate for 28% ($n = 402$) and low or very low for 45% ($n = 640$). Less than one quarter of reviews ($n = 63$) included at least one PIO with high-quality evidence that favoured a benefit of the experimental intervention evaluated in half of them ($n = 34$ reviews).

Conclusions

Many outcomes reported in the SoF table of recent Cochrane reviews can be considered PIOs. However, the quality of available evidence remains limited for these outcomes.

OPEN ACCESS

Citation: Yordanov Y, Dechartres A, Ravaud P (2018) Patient-important outcomes in systematic reviews: Poor quality of evidence. PLoS ONE 13(4): e0195460. <https://doi.org/10.1371/journal.pone.0195460>

Editor: Gianni Virgili, Università degli Studi di Firenze, ITALY

Received: November 14, 2017

Accepted: March 22, 2018

Published: April 5, 2018

Copyright: © 2018 Yordanov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data have been uploaded as Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In the last decade, clinicians and researchers have been encouraged to recommend healthcare interventions based on their effect on patient-important outcomes (PIOs) such as death, other clinical events, quality of life or functional outcomes [1–6]. However, several methodological reviews reported that the use of PIOs in randomized controlled trials (RCTs) of various medical specialities is still far from optimal [7–10]. In a recent report assessing trials of critically ill patients, less than one quarter of primary outcomes (24%) were PIOs [9]. This proportion was as low as 5% when exploring outcomes besides mortality after intensive care unit discharge, such as functional disability or quality of life [9]. Similar results were previously found for diabetes [8] or cardiovascular [7] trials, with 18% and 23% of these trials assessing PIOs. In contrast, some recent studies suggested that systematic reviews are more likely to evaluate PIOs than their individual trials [10, 11].

A key component of systematic reviews is the elaboration of a clear and focused clinical question by specifying the types of participants, interventions, comparisons, and outcomes that should be considered [12, 13]. When planning a Cochrane review, authors are strongly encouraged to consider all outcomes that are meaningful to patients, physicians, policy makers or any other health care stakeholders, regardless of their availability in individual trials [13, 14]. When reporting the review, they are asked to present a Summary of Findings (SoF) table summarizing the quality of evidence and treatment effect magnitude for the most essential outcomes for decision-making by using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [13–18].

In this study, we aimed to evaluate: 1) whether outcomes reported in the SoF table could be considered PIOs and 2) the quality of evidence for these outcomes in a sample of recent Cochrane reviews.

Methods

This is a methodological review of recently published Cochrane reviews. For each Cochrane review reporting an SoF table, we extracted all outcomes reported in this table and evaluated whether they could be considered PIOs. Then, we evaluated the quality of evidence for these outcomes.

Data sources

We previously obtained data from all systematic reviews published between March 2011 and September 2014 from the Cochrane Collaboration [19]. Data were provided as XML files and contained all information reported by the review authors using RevMan [20], the review software developed by the Cochrane Collaboration to prepare and update systematic reviews.

Review selection

Using R 3.2.2 and the XML package, we identified all reviews of RCTs including at least one SoF table. Reviews including observational studies were excluded. We also excluded reviews including only trials published before 2007 to focus on recent topics. Then, we manually examined all review titles and abstracts when necessary to select those evaluating a healthcare intervention (pharmacological or non-pharmacological). Reviews evaluating diagnostic test accuracy or economic evaluations were excluded. Among the set of eligible reviews, by using a random number generator, we drew a random sample of 300 reviews for comprehensive evaluation.

Identification of the SoF for the main comparison

Most reviews have a single SoF table, but some have several, corresponding to different comparisons. In this case, we manually identified the main comparison as reported by the review authors. If no main comparison was identified as such, we selected the one with the largest number of available outcomes and included the largest number of trials. When the same SoF table reported various comparisons, the review was excluded.

Data collection

For every included systematic review, we collected the following characteristics by using a standardized data extraction sheet:

- Review general characteristics: title, digital object identifier (DOI), protocol publication date, review first publication date, Cochrane Review Group (e.g., Gynaecological Cancer Group), type of intervention (i.e., pharmacological or non-pharmacological), number of excluded and included trials.
- Characteristics of the main comparison: Patient or population description, setting, intervention and comparator
- Every outcome reported in the SoF table including outcome description and follow-up period.

Classification of outcomes

We categorized each outcome as follows: mortality, other clinical events (e.g., myocardial infarction or stroke), therapeutic decision (e.g., transfusion), function (e.g., anxiety, depression, disability and dyspnoea), pain, quality of life, adverse events or side effects (identified as such by the review authors), physiological parameters (e.g., blood pressure, weight), biological parameters (e.g., cholesterol levels), radiological parameters (e.g., measure of joint space), compliance (e.g., discontinuation for any reason), process (e.g., duration of surgical procedure), resource use (hospitalisations), cost-effectiveness and satisfaction with care. A single reviewer classified all outcomes. As a quality measure, 10% of the outcomes were classified independently in duplicate (YY and AD).

Definition of PIOs

We considered mortality, other clinical events, adverse events, function, pain, quality of life and therapeutic decisions as PIOs consistent with previous studies [8, 11, 21, 22].

Quality of evidence

We evaluated the quality of evidence of each outcome reported in the SoF table based on the review authors assessment. The evidence was classified as high quality if further research is very unlikely to change the authors confidence in the estimate of effect; moderate quality if further research is likely to have an important impact on the authors' confidence in the estimate of effect and may change the estimate; low quality if further research is very likely to have an important impact on the confidence in the estimate of effect and likely to change the estimate; and very low quality if the estimate of the effect is very uncertain [18]. We evaluated separately the quality of evidence for outcomes that could be considered PIOs. For those considered with high quality of evidence, we also evaluated whether this was evidence of benefit for the

experimental intervention (results in favour of the experimental intervention: i.e., statistically significant results).

Statistical analysis

The analysis was mostly descriptive. Continuous data are presented as median (Q1–Q3) and qualitative data as frequency (percentage). All analyses involved use of R v3.1.1 (R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>).

Results

Selection of relevant Cochrane systematic reviews

The selection process was previously described[19]. Briefly, between March 2011 and September 2014, 2796 Cochrane systematic reviews were published, 1670 did not include an SoF table (60% of screened reviews) and 820 were eligible for inclusion. We identified a random sample of 300 reviews, but 10 were additionally excluded because their SoF presented different interventions within the same table, which left 290 reviews for further evaluation.

Characteristics of included reviews

The reviews were issued by 47 of the 53 Cochrane Review Groups[23]. Every review included a median of 11 trials overall (Q1–Q3: 5–21), with a median of 5 trials per main comparison (Q1–Q3: 3–12)[19]. The experimental intervention was non-pharmacological in 40% of the reviews (n = 115). Every SoF table included a median of 5 outcomes (Q1–Q3: 3–7), for a total of 1414 outcomes. The corresponding meta-analyses included a median of 3 trials (Q1–Q3: 2–7).

Outcome classification

Among the 1414 outcomes reported in the SoFs, the most common were functional outcomes (27%, n = 384), clinical events (14%, n = 198) and adverse events (12%, n = 174). Mortality represented 10% (n = 138) of the outcomes and quality of life 7% (n = 98). Among the outcomes, biological parameters represented 6% (n = 89), process and resource use 5% (n = 74), and physiological parameters and compliance 4% (n = 56) and 3% (n = 45). A total of 1089 outcomes (77%, 95% CI: 75–79) could be considered PIOs (Table 1). Almost all reviews included at least one PIO in the SoF table (n = 278, 96%; 95% CI: 93–98) (Fig 1). Per review, the median proportion of PIOs among the outcomes reported in the SoF was 86% (Q1–Q3: 67–100).

Quality of available evidence

The quality of evidence provided by the outcomes in the SoF tables was high for 12% (n = 168, 95% CI: 10–14) and moderate for 28% (n = 402, 95% CI: 26–31) (Table 2). For 45% of the outcomes (n = 640, 95% CI: 43–48), the quality of the evidence was low or very low. No GRADE assessment was available for 12% of the outcomes (n = 164, 95% CI: 10–13) because according to the review authors, these outcomes were not reported or measured in any individual trials. The quality of the available evidence was very close when focusing on only PIOs, with 41% (n = 449) having high and moderate evidence and 44% (n = 475) low to very low evidence.

Less than one-quarter of the reviews, 22% (n = 63; 95% CI: 17–27), included at least one PIO with evidence considered of high quality by the review authors. In half (n = 34) of these reviews, the results were in favour of the experimental intervention (i.e., statistically significant results) (Fig 1).

Table 1. Classification of outcomes reported in the Summary of Findings (SoF) table in 290 recent Cochrane reviews.

	No. (%)
	N = 1414 *
Patient-important outcomes (PIOs)	
Function	384 (27%)
Other clinical events	198 (14%)
Adverse events—side effects	174 (12%)
Mortality	138 (10%)
Quality of life	98 (7%)
Pain	71 (5%)
Therapeutic decision	33 (2%)
PIOs among the SoF table outcomes	1089 (77%; 95% CI 75–79)
Other outcomes	
Biological parameter	89 (6%)
Process, resource use	74 (5%)
Physiological parameters	56 (4%)
Compliance	45 (3%)
Satisfaction with care	24 (2%)
Radiological parameters	23 (2%)
Cost-effectiveness	16 (1%)

*The total is higher than N = 1414 because some outcomes were included in more than one category 95% CI, 95% confidence interval

<https://doi.org/10.1371/journal.pone.0195460.t001>

Discussion

In this study, we evaluated whether outcomes reported in the SoF table of recent Cochrane reviews could be considered PIOs and the quality of evidence provided for these outcomes. The large variety of review groups represented allowed us to explore very different medical specialties. More than three-quarters of the outcomes reported in the SoF tables could be considered PIOs. However, for a large proportion of the available evidence (45%), the quality was considered low or very low. About one quarter of the reviews included at least one PIO with high-quality evidence and for half of them, this evidence showed a benefit of the experimental intervention.

The Cochrane Collaboration strongly encourages review authors to include SoF tables to present their main findings [13]. These tables, first introduced in Cochrane reviews in 2008, aim to synthesize in a simple, transparent and accessible format, key information on the assessed interventions' magnitude of effect, sum of available data for the main outcomes and quality of evidence [13]. Therefore, SoF tables should include the most important outcomes whether they were measured in individual trials or not [13]. Previous studies reported that as compared with reviews without SoF tables, the inclusion of SoF tables seemed to improve readers' general understanding of the reviews and allowed them to better identify the critical information and find results for important outcomes (93% vs 44%) [24]. However, not all Cochrane reviews present a SoF table. In our study, we excluded 60% of Cochrane reviews because they had no SoF table, which is consistent with a previous study showing a proportion of inclusion of SoF tables evolving from 31% in 2008 to 57% in 2013 [25]. An alternative version of the SoF table called the interactive SoF (iSoF) allowing review authors to choose

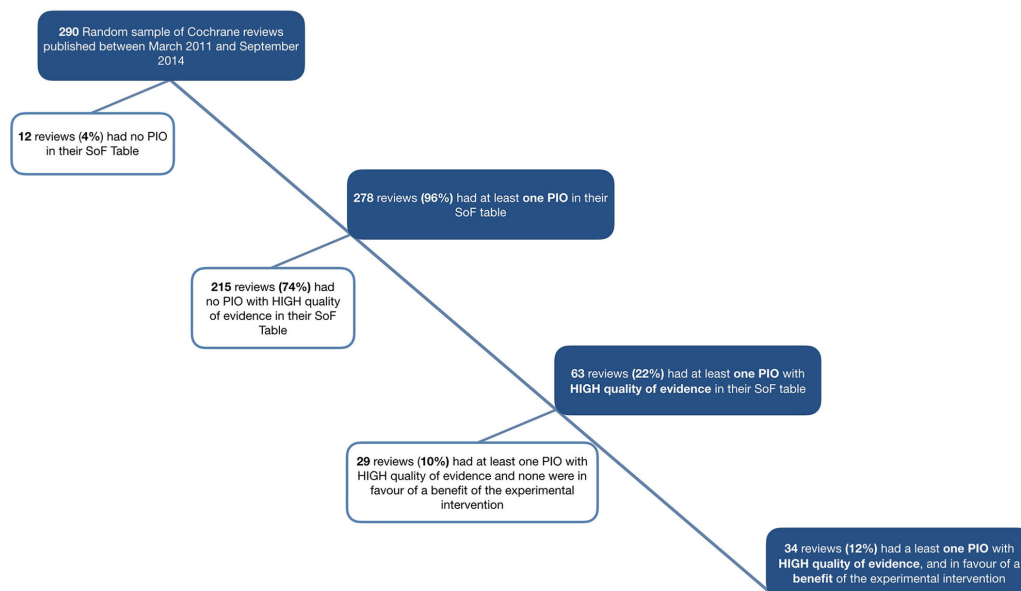


Fig 1. Patient-important outcomes in a random sample of Cochrane reviews.

<https://doi.org/10.1371/journal.pone.0195460.g001>

alternative displays (absolute effects rather than relative effects) is being tested[26], which may help generalise their use.

We found that most of the outcomes reported in the SoF tables could be considered PIOs, with a median proportion of PIOs in the SoF tables of 86%. In a study that focused on recently published Cochrane and non-Cochrane reviews but also on registered systematic reviews protocols (PROSPERO), Ameer *et al* described that 95% of Cochrane reviews included at least one PIO among their primary outcomes. This proportion was higher than for non-Cochrane reviews[11]. Two previous studies also found consistent results, with 68% to 71% of the reviews considering PIOs as primary outcomes[21, 22]. A large survey of the completeness of

Table 2. Quality of available evidence for outcomes reported in the SoF tables for the main comparison in 290 recent Cochrane reviews.

Quality of evidence	Proportion in %	All outcomes N = 1414	PIOs N = 1089
		Proportion in %	Proportion in %
High quality	Further research is very unlikely to change our confidence in the estimate of effect.	12	12
Moderate quality	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.	28	29
Low quality	Further research is very likely to have an important impact on our confidence in the estimate of effect and likely to change the estimate.	28	27
Very low quality	We are very uncertain about the estimate.	17	16
No GRADE	No individual trial reported or measured the desired outcome	12	12
No GRADE	Other reasons	3	3

<https://doi.org/10.1371/journal.pone.0195460.t002>

main outcomes mapped the availability of information for a major clinical outcome across randomized trials and systematic reviews in an entire discipline: 75% of reviews reported the major clinical outcome as compared with only 20% of primary trial reports [27]. All these results highlight the gap in outcomes between systematic reviews and their individual trials, which may explain why the quality of evidence was frequently limited for the outcomes reported in the SoF table, as shown by our results. In our study, only 22% of the reviews had at least one PIO with a high level of evidence, with half showing a benefit of the experimental intervention. The failure to consider PIOs in clinical trials may lead to erroneous evaluation of benefits, with possible serious consequences for patients, and represents waste [28–31]. To improve this situation, core outcome sets (COSs) should be developed (ie, standardized sets of outcomes, that have been agreed upon, and that should be measured and reported in every trial for a specific healthcare condition) and their use promoted [32–34]. The number of COSs developed is still limited but is progressively increasing, such as their use in systematic reviews [11, 25, 33, 35–37].

Our study has limitations. Although we relied on detailed and previously published definitions for PIOs, considerable judgment remains when assessing the importance of the outcomes, and patients or experts in the assessed medical field were not involved in this evaluation. Our results cannot be generalized to all systematic reviews because Cochrane reviews have precise conduct and reporting guidelines [13–15] and were described as more transparently reported than all other types of systematic reviews [38]. Our study confirms that many outcomes reported in the SoF tables of recent Cochrane reviews can be considered PIOs. However, the quality of evidence for these outcomes remains limited.

Conclusions

Our study confirms that many outcomes reported in the SoF tables of recent Cochrane reviews can be considered PIOs. However, the quality of evidence for these outcomes remains limited.

Supporting information

S1 Table. Quality of available evidence for outcomes reported in the SoF tables for the main comparison in 290 recent Cochrane reviews: Comparison of all outcomes, to PIOs and surrogate outcomes.

(DOCX)

S1 File. Supporting data 1: Complete list of outcomes, classified and categorized.

(XLSX)

S2 File. Supporting data 2: All the study data.

(XLSX)

Author Contributions

Conceptualization: Youri Yordanov, Agnes Dechartres, Philippe Ravaud.

Data curation: Youri Yordanov, Agnes Dechartres.

Formal analysis: Youri Yordanov, Agnes Dechartres, Philippe Ravaud.

Methodology: Youri Yordanov, Agnes Dechartres, Philippe Ravaud.

Resources: Youri Yordanov.

Supervision: Agnes Dechartres, Philippe Ravaud.

Validation: Agnes Dechartres.

Visualization: Youri Yordanov.

Writing – original draft: Youri Yordanov, Agnes Dechartres.

Writing – review & editing: Youri Yordanov, Agnes Dechartres, Philippe Ravaud.

References

1. Guyatt G, Montori V, Devereaux PJ, Schunemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP J Club*. 2004; 140(1):A11–2. PMID: [14711297](#).
2. Richards T, Montori VM, Godlee F, Lapsley P, Paul D. Let the patient revolution begin. *BMJ*. 2013; 346: f2614. Epub 2013/05/16. <https://doi.org/10.1136/bmj.f2614> PMID: [23674136](#).
3. Montori VM, Gandhi GY, Guyatt GH. Patient-important outcomes in diabetes—time for consensus. *Lancet*. 2007; 370(9593):1104–6. [https://doi.org/10.1016/S0140-6736\(07\)61489-5](https://doi.org/10.1016/S0140-6736(07)61489-5) PMID: [17905147](#).
4. Frank L, Basch E, Selby JV, Patient-Centered Outcomes Research I. The PCORI perspective on patient-centered outcomes research. *Jama*. 2014; 312(15):1513–4. <https://doi.org/10.1001/jama.2014.11100> PMID: [25167382](#).
5. Methodology Committee of the Patient-Centered Outcomes Research I. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *Jama*. 2012; 307(15):1636–40. <https://doi.org/10.1001/jama.2012.466> PMID: [22511692](#).
6. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *Jama*. 2012; 307(15):1583–4. <https://doi.org/10.1001/jama.2012.500> PMID: [22511682](#).
7. Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. *BMJ*. 2010; 341:c5707. <https://doi.org/10.1136/bmj.c5707> PMID: [21041324](#).
8. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. *Jama*. 2008; 299(21):2543–9. <https://doi.org/10.1001/jama.299.21.2543> PMID: [18523223](#).
9. Gaudry S, Messika J, Ricard JD, Guillo S, Pasquet B, Dubief E, et al. Patient-important outcomes in randomized controlled trials in critically ill patients: a systematic review. *Ann Intensive Care*. 2017; 7(1):28. <https://doi.org/10.1186/s13613-017-0243-z> PMID: [28271450](#).
10. Saldanha IJ, Li T, Yang C, Owczarzak J, Williamson PR, Dickersin K. Clinical trials and systematic reviews addressing similar interventions for the same condition do not consider similar outcomes to be important: a case study in HIV/AIDS. *J Clin Epidemiol*. 2017; 84:85–94. <https://doi.org/10.1016/j.jclinepi.2017.02.005> PMID: [28249722](#).
11. Ameur H, Ravaud P, Fayard F, Riveros C, Dechartres A. Systematic reviews of therapeutic interventions frequently consider patient-important outcomes. *J Clin Epidemiol*. 2017. <https://doi.org/10.1016/j.jclinepi.2017.01.006> PMID: [28147235](#).
12. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003; 96(3):118–21. PMID: [12612111](#).
13. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. www.cochrane-handbook.org.
14. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated September 2011]. ed. www.cochrane-handbook.org. The Cochrane Collaboration; 2011.
15. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Akl E, et al. Chapter 11: Completing 'Summary of findings' tables and grading the confidence in or quality of the evidence. In: Higgins JPT, Chandler J, (editors) CM, editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 5.2.0 (updated June 2017). ed. Cochrane, 2017. <http://training.cochrane.org/handbook>. The Cochrane Collaboration; 2017.
16. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ. What is "quality of evidence" and why is it important to clinicians? *Bmj*. 2008; 336(7651):995–8. Epub 2008/05/06. <https://doi.org/10.1136/bmj.39490.551019.BE> PMID: [18456631](#).
17. 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. Epub 2004/06/19. <https://doi.org/10.1136/bmj.328.7454.1490> PMID: [15205295](#).

18. 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. <https://doi.org/10.1136/bmj.39489.470347.AD> PMID: 18436948.
19. Yordanov Y, Dechartres A, Atal I, Tran V, Boutron I, Crequit P, et al. Avoidable waste of research related to outcome planning and reporting in clinical trials. *BMC Medicine—under review*. 2017.
20. Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. *BMC medicine*. 2016; 14(1):100. <https://doi.org/10.1186/s12916-016-0639-x> PMID: 27377062.
21. Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, Akl EA, Vernooij RW, et al. Systematic reviews experience major limitations in reporting absolute effects. *J Clin Epidemiol*. 2016; 72:16–26. <https://doi.org/10.1016/j.jclinepi.2015.11.002> PMID: 26560992.
22. Agarwal A, Johnston BC, Vernooij RW, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, et al. Authors seldom report the most patient-important outcomes and absolute effect measures in systematic review abstracts. *J Clin Epidemiol*. 2016. <https://doi.org/10.1016/j.jclinepi.2016.08.004> PMID: 27555080.
23. Cochrane Collaboration. "Cochrane Review Groups"[Internet] 2017 [Date Accessed: 19/09/2017] [cited 2017 19/09/2017]. www.cochrane.org/contact/review-groups.
24. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol*. 2010; 63(6):620–6. <https://doi.org/10.1016/j.jclinepi.2009.12.014> PMID: 20434024.
25. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev*. 2016; 5:11. <https://doi.org/10.1186/s13643-016-0188-6> PMID: 26792080.
26. Alonso-Coello P, Vernooij RWM, Carrasco-Labra A, Vandvik P, Guyatt GH. RE: Which estimates comprise optimal reporting in systematic reviews? *Journal of Clinical Epidemiology*. 2017; 81(Supplement C):140–1. <https://doi.org/10.1016/j.jclinepi.2017.01.001>.
27. Ioannidis JP, Horbar JD, Ovelman CM, Brosseau Y, Thorlund K, Buus-Frank ME, et al. Completeness of main outcomes across randomized trials in entire discipline: survey of chronic lung disease outcomes in preterm infants. *BMJ*. 2015; 350:h72. Epub 2015/01/28. <https://doi.org/10.1136/bmj.h72> PMID: 25623087.
28. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009; 374(9683):86–9. [https://doi.org/10.1016/S0140-6736\(09\)60329-9](https://doi.org/10.1016/S0140-6736(09)60329-9) PMID: 19525005
29. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014; 383(9912):166–75. [https://doi.org/10.1016/S0140-6736\(13\)62227-8](https://doi.org/10.1016/S0140-6736(13)62227-8) PMID: 24411645.
30. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014; 383(9912):101–4. [https://doi.org/10.1016/S0140-6736\(13\)62329-6](https://doi.org/10.1016/S0140-6736(13)62329-6) PMID: 24411643.
31. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014; 383(9913):267–76. Epub 2014/01/15. [https://doi.org/10.1016/S0140-6736\(13\)62228-X](https://doi.org/10.1016/S0140-6736(13)62228-X) PMID: 24411647.
32. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy*. 2012; 17(1):1–2. <https://doi.org/10.1258/jhsrp.2011.011131> PMID: 22294719
33. Williamson P, Clarke M. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. *Cochrane Database Syst Rev*. 2012; 5:ED000041. Epub 2012/05/18. <https://doi.org/10.1002/14651858.ed000041> PMID: 22592744.
34. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials*. 2012; 13:132. Epub 2012/08/08. <https://doi.org/10.1186/1745-6215-13-132> PMID: 22867278
35. Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews?—a survey of the Co-ordinating Editors of Cochrane Review Groups. *Trials*. 2013; 14. <https://doi.org/10.1186/1745-6215-14-21> PMID: 23339751
36. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev*. 2016; 5(1):11. <https://doi.org/10.1186/s13643-016-0188-6> PMID: 26792080.
37. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. *PLoS ONE*. 2016; 11(1):e0146444. <https://doi.org/10.1371/journal.pone.0146444> PMID: 26785121.
38. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS Med*. 2016; 13(5):e1002028. <https://doi.org/10.1371/journal.pmed.1002028> PMID: 27218655.

Supporting table 1. Quality of available evidence for outcomes reported in the SoF tables for the main comparison in 290 recent Cochrane reviews

Quality of evidence	Definition	All outcomes	PIOs	Surrogate outcomes
		N =1414 Proportion in %	N =1089 Proportion in %	N =168 Proportion in %
High quality	Further research is very unlikely to change our confidence in the estimate of effect.	12%	12%	12%
Moderate quality	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.	28%	29%	30%
Low quality	Further research is very likely to have an important impact on our confidence in the estimate of effect and likely to change the estimate.	28%	27%	38%
Very low quality	We are very uncertain about the estimate.	17%	16%	14%
No GRADE	No individual trial reported or measured the desired outcome	12%	12%	5%
No GRADE	Other reasons	3%	3%	1%

RESEARCH ARTICLE

Open Access



Avoidable waste of research related to outcome planning and reporting in clinical trials

Youri Yordanov^{1,2,3,4*} , Agnes Dechartres^{1,4,5,6}, Ignacio Atal^{1,4}, Viet-Thi Tran^{1,4}, Isabelle Boutron^{1,4,5,6}, Perrine Crequit^{1,4} and Philippe Ravaud^{1,4,5,6,7}

Abstract

Background: Inadequate planning, selective reporting, and incomplete reporting of outcomes in randomized controlled trials (RCTs) contribute to the problem of waste of research. We aimed to describe such a waste and to examine to what extent this waste could be avoided.

Methods: This research-on-research study was based on RCTs included in Cochrane reviews with a summary of findings (SoF) table. We considered the outcomes reported in the SoF tables as surrogates for important outcomes for patients and other decision makers. We used a three-step approach. (1) First, in each review, we identified, for each important outcome, RCTs that were excluded from the corresponding meta-analysis. (2) Then, for these RCTs, we systematically searched for registrations and protocols to distinguish between inadequate planning (an important outcome was not reported in registries or protocols), selective reporting (an important outcome was reported in registries or protocols but not in publications), and incomplete reporting (an important outcome was incompletely reported in publications). (3) Finally, we assessed, with the consensus of five experts, the feasibility and cost of measuring the important outcomes that were not planned. We considered inadequately planned or selectively or incompletely reported important outcomes as avoidable waste if the outcome could have been easily measured at no additional cost based on expert evaluation.

Results: Of the 2711 RCTs included in the main comparison of 290 reviews, 2115 (78%) were excluded from at least one meta-analysis of important outcomes. Every trial contributed to 55%, on average, of the meta-analyses of important outcomes. Of the 310 RCTs published in 2010 or later, 156 were registered. Inadequate planning affected 79% of these RCTs, whereas incomplete and selective reporting affected 41% and 15%, respectively. For 63% of RCTs, we found at least one missing important outcome for which the waste was avoidable and for 30%, the waste was avoidable for all important outcomes.

Conclusions: Most of the RCTs included in our sample did not contribute to all the important outcomes in meta-analyses, mostly because of inadequate planning or incomplete reporting. A large part of this waste of research seemed to be avoidable.

Keywords: Randomized controlled trial, Waste of research, Outcome, Selective reporting, Core outcome set

* Correspondence: youri.yordanov@aphp.fr

¹INSERM, U1153, Hôpital Hôtel-Dieu, 1, place du parvis Notre Dame, 75004 Paris, France

²Sorbonne Universités, UPMC Paris Univ-06, Paris, France

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Clinical trials are only as credible as their outcomes, so to inform decision-making appropriately, randomized controlled trials (RCTs) must evaluate the outcomes that are most important to patients and their caregivers [1–8]. Failure to do so could contribute to the overwhelming problem of waste in research [9–16]. Waste of research related to inadequate outcome planning, selective reporting, and incomplete reporting of outcomes in RCTs prevents patients and their physicians from making well-informed decisions, with potential serious consequences if ineffective or harmful treatments are promoted [9, 12–14, 17, 18].

Therefore, when planning an RCT, researchers are expected to measure all important outcomes [9]. However, previous studies found that less than one-fifth of diabetes RCTs and less than one-quarter of cardiovascular trials considered patient-important outcomes as their primary outcomes [19, 20]. Instead, researchers frequently rely on surrogates as a proxy for final patient-important outcomes because these outcomes allow for smaller, faster, and thus, cheaper clinical trials [21–24]. However, surrogates can be misleading, because they may show exaggerated treatment effect sizes or even an apparent benefit of harmful treatments, as was the case for the use of antiarrhythmic drugs after myocardial infarction, which led to the deaths of several thousand patients decades ago [25–28].

Selective reporting has been repeatedly described as another important issue affecting RCT outcomes [29–32]. Outcome reporting bias arises when outcomes are selectively reported based on the nature and direction of the results [33]. In a recent study, the median proportion of RCTs with discrepancies between registered and published primary outcomes was 31% [34]. Statistically significant outcomes were 2 to 4 times more likely to be reported in publications than non-significant ones, which biases the available body of evidence toward more positive results [30].

Similarly, a median of 31% to 50% of efficacy outcomes were found to be incompletely reported in RCT articles [29, 35, 36]. With incomplete reporting of outcomes, outcomes cannot be included in meta-analyses, which poses a serious threat to the usability of trial results: not including all available results in meta-analyses can lead to a truncated vision of the overall body of evidence.

The purpose of this study was to describe the waste of research related to inadequate planning, selective reporting, or incomplete reporting of outcomes in RCTs and to examine to what extent this waste could be avoided. For this, we addressed the following specific questions: (1) What proportion of RCTs are excluded from meta-analyses due to outcome reasons? (2) Were the exclusions related to inadequate planning, selective reporting, or incomplete reporting? (3) Was it feasible to

measure the missing outcomes at the planning stage, and at what cost?

Methods

We performed a research-on-research study based on RCTs included in Cochrane systematic reviews. We used the outcomes reported in the summary of findings (SoF) tables of reviews as surrogates for important outcomes, because Cochrane systematic review SoF tables should include the most important outcomes for patients and other decision makers, whether they are available in RCTs or not [7]. This study used a three-step approach. (1) First, for all important outcomes, we identified the RCTs that were included in the Cochrane reviews but excluded from the corresponding meta-analyses because the important outcome was missing. (2) Then, we systematically searched for trial registrations and protocols to distinguish between the outcomes that were not planned (not included in registries or protocols, i.e., inadequate planning) and outcomes that were planned. For the planned outcomes, we distinguished between those that were adequately reported and those that were incompletely reported (poor reporting) or not reported (selective reporting). (3) Finally, we assessed, via expert consensus, the feasibility and cost of measuring the outcomes that were not planned.

The eligibility criteria at each phase of the study are summarized in Additional file 1.

Identification of RCTs excluded from meta-analyses

Data sources

We obtained data from all Cochrane systematic reviews published between March 2011 and September 2014. They were provided by the Cochrane Collaboration editorial unit as XML files and contained all information reported by the review authors in RevMan, the software developed by the Cochrane Collaboration for preparing and maintaining systematic reviews [37]. Cochrane systematic reviews of interventions are organized by comparisons of two treatment groups. Meta-analyses are then performed for one or more outcomes within each comparison. Cochrane reviewers are encouraged to present an SoF table summarizing information on the quality of evidence and treatment effect magnitude (from the meta-analysis result) for the most important outcomes (see example in the review by Mocellin and colleagues [38]) [7]. According to the Cochrane collaboration, these outcomes should be important to all research end users, that is, patients and other decision makers [7]. These outcomes include a “wide variety of events such as mortality and major morbidity (such as stroke and myocardial infarction); however, they may also represent frequent minor and rare major side effects, symptoms and quality of life, burdens associated with treatment, and resource issues (costs)” [7].

Selection of relevant Cochrane systematic reviews

Using R 3.1.1 and the XML package, one of us (IA) removed from the set of reviews provided by the Cochrane Collaboration withdrawn Cochrane reviews and then identified all reviews of RCTs with an SoF table. We excluded reviews that included observational studies and those including only RCTs published before 2007 because we focused on recent RCTs. From all eligible Cochrane reviews, using a random number generator, we drew a random sample of 300 for an in-depth evaluation but excluded a further 10 reviews because their SoF tables mixed various interventions, which resulted in a final sample of 290 Cochrane reviews.

Identification of important outcomes

We considered outcomes reported in the SoF table as a surrogate for important outcomes. Therefore, we manually identified and extracted all outcomes reported in the SoF tables. Most reviews had a single SoF table, but some had several tables, corresponding to different comparisons. In this case, we focused on the main comparison as acknowledged by the authors. If the main comparison was not reported by the review authors or if various comparisons were reported in the same SoF table, we selected the comparison with the most outcomes and included the largest number of RCTs. If the SoF table reported various interventions (e.g., presented three meta-analyses of different outcomes for three different interventions), the review was excluded.

We classified each outcome as follows: mortality, other clinical event (e.g., myocardial infarction or stroke), therapeutic decision (e.g., transfusion), function (e.g., disability), pain, quality of life, adverse events or side effects (identified as such by the review authors), physiological variable (e.g., blood pressure or weight), biological variable (e.g., cholesterol levels), radiological variable (e.g., measure of joint space), compliance (e.g., discontinuation for any reason), process (e.g., duration of surgical procedure), resource use (hospitalization), cost-effectiveness, and satisfaction with care [39].

Identification of RCTs excluded from meta-analyses

For each review and for each important outcome, we identified all RCTs excluded from the outcome in the corresponding meta-analysis. To do so, by using the reference list of all studies included in the Cochrane review, we first identified all RCTs available for the selected comparison by extracting all RCTs included in any meta-analysis reported for this comparison. Then, for each of these RCTs, we manually evaluated whether they contributed to the meta-analysis of each important outcome by screening RCTs included in the corresponding meta-analysis.

Evaluation of the reason for a missing outcome

We a priori hypothesized that the exclusion of an RCT from a meta-analysis for outcome reasons could be related to inadequate planning (i.e., the outcome was not planned to be measured), selective reporting (i.e., the outcome was planned but not reported in the publication), or incomplete reporting (i.e., the outcome was reported in the publication but not in a way that allowed for pooling of data).

To distinguish among inadequate planning, selective reporting, or incomplete reporting, for each RCT excluded from at least one meta-analysis, we screened the data available in Cochrane reviews and systematically searched for trial registration and/or protocols.

We focused on RCTs published in 2010 or later to maximize the chance of identifying trial registration or protocols.

Search for RCT registration and protocols

For each RCT, we individually assessed all available reports and information. To do that, we extracted the references identified from the “References to studies included in this review” section of the Cochrane reviews. One of us (YY) retrieved all the articles, conference abstracts, reports etc. related to the identified RCT. We screened all articles for any information regarding registration (name of trial registry and/or registration number) and/or protocol availability. When no reference to a trial registration was found, we used the following approach:

1. We searched the full text of the Cochrane review for any information regarding trial registration.
2. If no information was found, we searched the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP). Also, according to the author's affiliation, we searched the local registry if not part of the WHO ICTRP. We used the article title, the first and last author's name as author or as investigator, or the author's affiliation as search keywords.
3. If no information was found, we searched Google with the publication title and keywords regarding registration (e.g., registry, registration, NCT etc.).
4. If no information was found, we contacted the corresponding author to ask whether the trial was registered and whether a protocol was available.

Evaluation of the reason for the missing outcome

For each identified RCT, one of us (YY) classified each missing outcome into one of the following five categories:

1. Inadequate planning
 - The outcome was not planned according to the protocol or the registry entry. It was not reported in the available trial reports.

2. Selective reporting
 - The outcome was planned according to the protocol or the registry entry but was not reported in the available trial reports.
3. Incomplete reporting
 - The outcome was planned according to the protocol or the registry entry and was reported in the available trial reports but incompletely or not in a way that allowed for meta-analysis (e.g., mixed model analyses reported, no results per group, missing control group results, difference in means without the standard error, etc.).
 - No protocol or registry entry was found. The outcome was reported in the available trial reports but not in a way that allowed for meta-analysis.
4. Unable to distinguish between selective reporting and inadequate planning
 - No protocol or registry entry was found and the outcome was not reported in the available trial reports.
5. Other situations
 - The outcome was listed in the trial reports, but there was no event (e.g., the outcome was death but no death occurred).
 - The outcome concerned adverse events, but there was no event (e.g., the outcome was “Major Complications—Visceral injury,” but no complications were reported).
 - The outcome was reported in the available trial reports in a way that could allow for inclusion in a meta-analysis but was not included in the meta-analysis.

As a quality assessment measure, 10% of the RCTs were classified independently by two reviewers (YY and AD), with no disagreements.

Evaluation of research waste

Feasibility and cost of measuring the missing outcomes that were not planned

To evaluate whether the missing outcomes that were not planned could have been easily measured, we used an expert consensus approach. One of us (YY) presented to an expert panel of five methodologists and trialists (AD, IB, PC, PR, and VT) each RCT for which at least one missing outcome was classified as not planned. After evaluating standardized information on the population, interventions in the experimental and control group, and other outcomes evaluated, the panel of experts were asked to answer the following questions regarding the missing outcomes:

- “According to you, given the other outcomes measured in the trial, and based on your experience, would you consider that measuring the presented outcome would be easy, moderately easy, difficult or impossible in most cases from a trialist’s perspective?”
 - “According to you, given the other outcomes measured in the trial, and based on your experience, would you consider that measuring the presented outcome would be easy, moderately easy, difficult or impossible in most cases from a patient’s perspective?”
 - “According to you, given the other outcomes measured in the trial, and based on your experience, what would be the approximate cost of measuring this outcome: no cost defined as $\leq 1\%$ of the total cost of the trial; minor cost, defined as $\leq 5\%$; moderate cost, 5% to 15%; or major cost, 15% or more”. These percentages were indicative.
 - “According to you, how important is the missing outcome: major importance; non-major importance?”
- One of us (YY) ensured that all experts first gave their opinion and then discussed together, to avoid the opinion of one leading person influencing the others. The final feasibility and cost evaluation of every proposed adjustment was based on the group consensus.
- The qualifications and areas of expertise of the experts involved are summarized in Additional file 2.

Avoidable waste of research related to missing outcomes

We defined avoidable waste related to missing outcomes as missing important outcomes related to:

- Selective reporting
- Incomplete reporting
- Inadequate planning, if the outcome was judged by the expert panel as easy to collect from both the trialist and patient perspective, not costly (i.e., no or minor additional cost), and of critical importance

At the trial level, such waste could have been partially avoided if the trial could have been included in the meta-analysis of at least one missing outcome and totally if it could have been included in all meta-analyses of missing outcomes.

Statistical analysis

The analysis was mainly descriptive. Continuous data are presented as median (Q1–Q3) and categorical data as frequencies (%), 95% confidence interval [CI]. All analyses involved use of R 3.0.2 (2013–09-25) (R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/>).

Results

Identification of RCTs excluded from meta-analyses

Selection and characteristics of the Cochrane systematic reviews

The complete selection process is described in Fig. 1. Briefly, from the 2796 Cochrane systematic reviews published between March 2011 and September 2014, 820 reviews corresponded to our eligibility criteria. Our selection process resulted in a sample of 290 reviews (5047 RCTs), with a median of 11 RCTs per review (Q1–Q3: 5–21) (Fig. 1). The subset of included reviews appeared comparable to the eligible Cochrane Reviews (Additional file 3). The reviews investigated 47 different health research topics, the most common being airways (7%, $n = 21/290$), menstrual disorders and subfertility (7%, $n = 20/290$), and anesthesia (6%, $n = 18/290$) (Table 1). Experimental interventions were pharmacological in 60% of reviews ($n = 175/290$).

The reviews included a median of 5 outcomes per SoF table for the main comparison (Q1–Q3: 3–7, Min–Max: 1–12), for a total of 1414 outcomes. Mortality represented 10% ($n = 138/1414$) of the outcomes, and other clinical events, 14% ($n = 198/1414$); quality of life, 7% ($n = 98/1414$); function, 27% ($n = 384/1414$); and adverse events, 12% ($n = 174/1414$) (Table 1). Biological variables, process and resource use, and physiological variables accounted for 6% ($n = 89/1414$), 5% ($n = 74/1414$), and 4% ($n = 56/1414$), respectively. The corresponding meta-analyses included a median of 3 RCTs (Q1–Q3: 2–7), maximum 76.

Proportion of RCTs excluded from the meta-analyses due to outcome reasons

The 290 reviews included a total of 2711 RCTs in the main comparison; 596 (22%; 95% CI: 20–24) were included in all meta-analyses of important outcomes, and

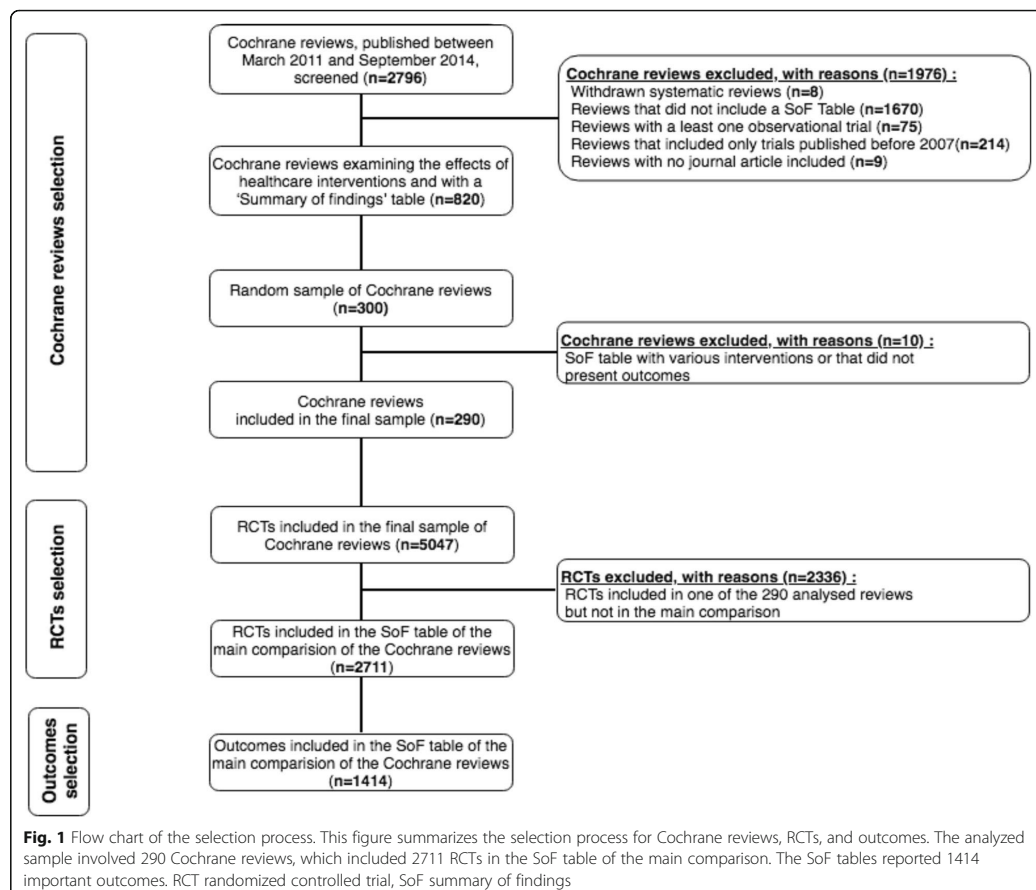


Fig. 1 Flow chart of the selection process. This figure summarizes the selection process for Cochrane reviews, RCTs, and outcomes. The analyzed sample involved 290 Cochrane reviews, which included 2711 RCTs in the SoF table of the main comparison. The SoF tables reported 1414 important outcomes. RCT randomized controlled trial, SoF summary of findings

Table 1 Characteristics of selected systematic reviews and outcomes

Review characteristics	<i>n</i> = 290
	Median (Q1–Q3)
No. of trials per review	11 (5–21)
No. of trials in the main comparison	5 (3–12)
No. of comparisons per review	2 (1–5)
Cochrane review groups	<i>n</i> = 290
	No. (%)
Airways	21 (7)
Menstrual disorders and subfertility	20 (7)
Anesthesia	18 (6)
Oral health	15 (5)
Schizophrenia	15 (5)
Hepato-biliary	12 (4)
Neuromuscular disease	12 (4)
Musculoskeletal	11 (4)
Infectious diseases	10 (3)
Other	156 (54)
Type of assessed intervention	<i>n</i> = 290
	No. (%)
Pharmacological	175 (60)
Non-pharmacological	115 (40)
Outcomes characteristics	<i>n</i> = 141
No. of outcomes per SoF table, median (Q1–Q3)	5 (3–7)
No. of trials per meta-analysis, median (Q1–Q3)	3 (2–7)
Outcome categories	<i>n</i> = 1414
	No. (%) [*]
Function	384 (27)
Other clinical events	198 (14)
Adverse events; side effects	174 (12)
Mortality	138 (10)
Quality of life	98 (7)
Biological variables	89 (6)
Process, resource use	74 (5)
Pain	71 (5)
Physiological variables	56 (4)
Compliance	45 (3)
Therapeutic decision	33 (2)
Satisfaction with care	24 (2)
Radiological variables	23 (2)
Cost-effectiveness	16 (1)

^{*}The total exceeds 100% because some outcomes were included in more than one category

2115 (78%, 95% CI: 76–80) were excluded from at least one meta-analysis. Every RCT contributed to 55%, on average, of the meta-analyses for important outcomes.

Evaluation of the reason for missing outcomes

Among the 2115 RCTs excluded from at least one meta-analysis of important outcomes, 310 were published in 2010 or later. We further excluded 19 RCTs, because their reports were not accessible (e.g., retracted paper) or were not in English or French (e.g., Chinese), which left 291 RCTs for further evaluation, with a total of 971 missing outcomes. A registration number or a protocol was retrieved for 54% (*n* = 156/291) of these RCTs (corresponding to 461 missing outcomes).

Reasons for missing outcomes were incomplete reporting for 21% of missing outcomes (*n* = 204/971) in 40% of RCTs (*n* = 117/291), and inadequate planning for 29% of missing outcomes (*n* = 282/971) in 42% of RCTs (*n* = 123/291) (Table 2). Confirmed selective reporting represented 4% of missing outcomes (*n* = 36/971) in 9% of RCTs (*n* = 25/291). Nevertheless, 42% of RCTs (*n* = 122/291) had no protocol or registration, so for these, we could not distinguish between selective reporting and inadequate planning.

When restricting our description to registered RCTs (*n* = 156), inadequate planning concerned 61% of missing outcomes (*n* = 282/461) in 79% of RCTs (*n* = 123/156), whereas incomplete reporting accounted for 21% of missing outcomes (*n* = 98/461) in 41% of RCTs (*n* = 64/156), and selective reporting, 7% of missing outcomes (*n* = 34/461) in 15% of RCTs (*n* = 23/156) (Table 2).

Evaluation of research waste

Feasibility of measuring the missing outcomes that were not planned

We submitted the 282 outcomes missing due to inadequate planning from 123 RCTs to our panel of experts. For 78% of outcomes (*n* = 221/282), the experts judged the outcome of critical importance given the context. For these 221 critically important outcomes, they considered that 82% (*n* = 182) could have been easily measured from both the trialist and patient perspective at no cost (Additional file 4).

Avoidable waste of research due to missing important outcomes

For the 291 RCTs published in 2010 or later, taking into account selective or incomplete outcome reporting, waste of research could have been partially avoided for 43% (*n* = 126) and totally (i.e., the trial could have been included in all meta-analyses of important outcomes) for 12% (*n* = 34) (Fig. 2). If we also consider missing outcomes that could have been easily measured from the planning stage at no additional cost as judged by our experts, waste of research could have been partially avoided for 63% of RCTs (*n* = 183) and totally for 30% (*n* = 86) (Fig. 2).

Table 2 Classification of reasons for missing outcomes

Reasons for missing outcome	All trials published in 2010 or later			
	All trials (registered and unregistered)		Registered trials	
	No. of affected outcomes (%) N = 971	No. of affected trials (%) ^a N = 291	No. of affected outcomes (%) N = 461	No. of affected trials (%) ^a N = 156
Inadequate planning	282 (29)	123 (42)	282 (61)	123 (79)
Selective reporting	36 (4)	25 (9)	34 (7)	23 (15)
Incomplete reporting	204 (21)	117 (40)	98 (21)	64 (41)
Unable to distinguish between selective reporting and lack of planning	363 (39)	122 (42)		
Other situations	86 (9)	63 (24)	47 (10)	41 (26)

^aThe total exceeds 100% because some outcomes were included in more than one category

Discussion

We evaluated the avoidable waste of research due to outcome-related reasons across a large number of RCTs included in recent Cochrane systematic reviews in a variety of medical fields. Our analysis revealed that 78% of the RCTs were not included in all meta-analyses of important outcomes and that every RCT contributed to 55%, on average, of the meta-analyses of these important outcomes. Among registered RCTs, inadequate planning was the most common reason affecting 79% of RCTs with missing outcomes. Such waste could have been partially avoided for 63% of the RCTs and totally for 30%.

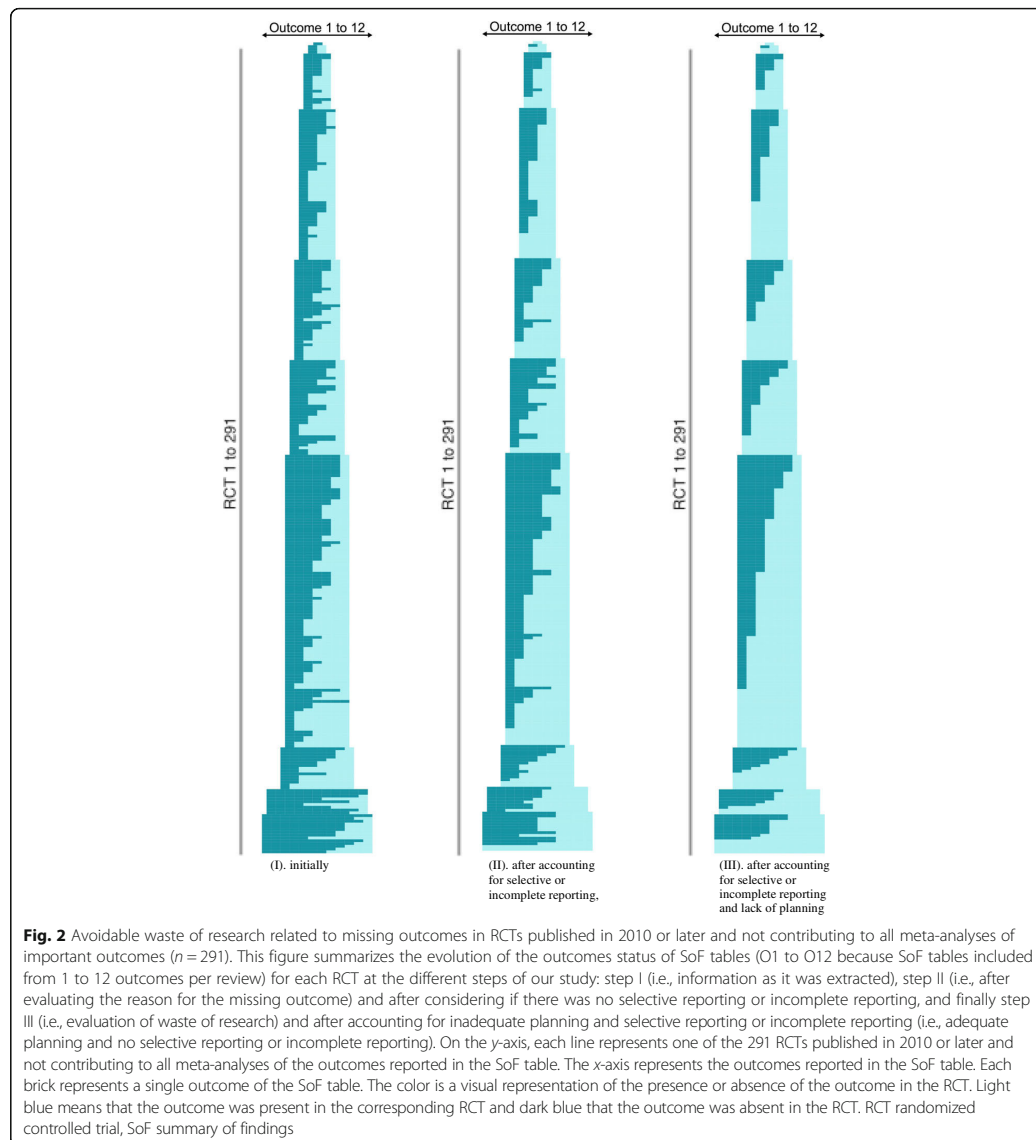
Our findings suggest that many RCTs were not included in all meta-analyses of important outcomes and that every RCT contributed to only half of the meta-analyses, on average. These results seem consistent with the literature. In a recent study of patient-important outcomes, the median proportion of trials included in a meta-analysis for such outcomes was about 60% [39]. Approximately half of the RCTs included in another sample of Cochrane reviews were included in the pooled effect size estimates in the meta-analyses of patient-important outcomes [40], whereas more than one-third of the outcomes prespecified in the review were not reported in the results sections [18, 41].

Although possible reasons for excluding RCTs from meta-analyses have been hypothesized, the respective contribution of inadequate planning and selective reporting or incomplete reporting was unknown. Our results provide major insights into this question, showing that many RCTs do not plan the most important outcomes to evaluate. Inadequate planning represents a missed opportunity within a trial because equipoise remains unchanged after this trial. However, a missed opportunity at the trial level will result in waste of research at the meta-analysis level. Actually, clinical research should be considered a sequential process, with each new trial contributing to the existing body of evidence. If a trial fails to do so because of a missed opportunity, it becomes a source of waste of research and resources

at the meta-analysis level because this trial will not contribute to the overall evidence. Cochrane reviewers may select outcomes that differ in importance from those selected by trialists (typically those defined as primary outcomes and used for sample size calculation). Trialists rarely consider adverse events (particularly severe adverse events) as primary outcomes and for sample size calculation. However, these outcomes are crucial to assess and report because they could be included in meta-analyses to increase power and inform decision-making. Identifying important outcomes is challenging, and without a consensus on outcomes, heterogeneity in outcomes evaluated will remain [41].

To overcome these issues, several initiatives have emerged to improve the relevance and consistency of outcomes used in clinical research [4, 42–45]; one is the Core Outcome Measures in Effectiveness Trials (COMET) initiative promoting the development and use of standardized core outcome sets (COSs). COMET offers extremely helpful tools for researchers aiming to develop COSs in their fields. However, although the number of COSs is rapidly increasing, it remains limited [46–48], and searching for specific existing COSs for a given condition is difficult. The COMET website may evolve to present available COSs in a tabular fashion to help trialists identify the outcomes to use and researchers the conditions for which COSs are needed. Another perspective would be to use the outcomes reported in the SoF table of Cochrane reviews to develop COSs, because most of these outcomes seem important to patients.

Incomplete reporting was another common reason for outcomes excluded from meta-analyses. Two main situations can be distinguished. First, the failure to report results adequately (e.g., reporting means without standard deviations) and second, the reporting of an analysis that cannot be included as such in a meta-analysis (e.g., reporting of repeated data analysis). The first situation is clearly related to poor reporting. We previously showed that results were more completely reported at



ClinicalTrials.gov than in publications, probably because of the standardized template used to report results at ClinicalTrials.gov [49]. Such standardized templates may help trialists determine which data should be reported and should be included in reporting guidelines and required by journals to improve the presentation of results. The other situation is more complex because it does not reflect poor

reporting, just reporting in a different way. Studies with repeated data are meant to record measurements at numerous time points to inform researchers of changes over time. However, several approaches for meta-analyzing these types of data exist, and they can differ in terms of the data needed for analysis [50]. Therefore, reporting trial results not just for an immediate use but also considering a later inclusion in

meta-analyses can be challenging. The Instrument for reporting of Planned Endpoints in Clinical Trials (InsPECT) reporting guidelines (currently under development, <https://www.inspect-statement.org/>) might help trial authors mitigate the effects of incomplete reporting in RCTs.

Finally, selective reporting of outcomes seems to affect a few trials. It remains an important issue in clinical trials, with, on average, one-third of discrepancies in primary outcomes between protocols and publications [29] or between registry information and publications [31]. In our situation, this concerns only outcomes that were planned and not reported at all in publications. Actually, reviewers consider a given outcome whether this outcome is reported as a primary or a secondary outcome.

Strengths and limitations of the study

In this study, we used an original approach to evaluate the main reasons why a trial was not included in a meta-analysis and the part of this waste that could be avoided. We analyzed a large set of trials included in a vast, unselected, sample of recent Cochrane reviews exploring various health-care research topics. Assembling numerous experts in such a variety of fields to identify the most important outcomes to evaluate would have been challenging. As a proxy, we used the outcomes reported in the SoF tables that were considered by the review authors as the most important to measure for a given comparison in a particular health condition [7]. We used the consensus of an expert panel of recognized methodologists and trialists to assess the feasibility and costs of measuring the missing outcomes. To avoid over-estimating waste of research, we also asked the panel to evaluate the importance of outcomes and considered that research was wasted for only those outcomes the panel confirmed to be of critical importance.

Our study has several limitations. First, we focused only on trials included in Cochrane reviews published between 2010 and 2014. However, it has been reported that Cochrane and non-Cochrane reviews have systemic differences, likely reflective of different methodology [51]. Then, we focused on trials included in the reviews but excluded from at least one meta-analysis. We did not consider the number of trials excluded from the reviews for outcome-related reasons. Therefore, we probably underestimated the proportion of trials excluded from meta-analyses for outcome-related reasons. Despite the recommendation of the Cochrane Handbook to determine trial eligibility independently of outcomes measured (studies should not be excluded just because they provide no usable data), some trials may be excluded for this reason [7]. Thus, although, our sample of RCTs was from a large variety of medical fields, the database used (Cochrane reviews published between 2010 and 2014

and with an SoF table), the limited sample size, the language restrictions, or the expertise of our panel of experts may limit the generalizability of our results. Finally, as a quality measure, 10% of the RCTs were classified independently by two reviewers, since this assessment is challenging and implies some subjectivity.

Conclusions

Our study shows that most RCTs included in our sample of Cochrane reviews did not contribute to all meta-analyses of the most important outcomes mainly because of inadequate planning or incomplete reporting. Such waste could have been partially avoided for 63% of the trials and totally for 30%. We need to accelerate the development and dissemination of COSs and reduce poor outcome reporting to avoid this waste.

Additional files

Additional file 1: Summary of the inclusion and exclusion criteria for the different phases of the study. (DOCX 18 kb)

Additional file 2: Qualifications and areas of expertise of the experts involved in the final step of the study. (DOCX 15 kb)

Additional file 3: Characteristics of the 820 Cochrane systematic reviews and the analyzed subset of 290 reviews. (DOCX 15 kb)

Additional file 4: Experts' opinion of the feasibility and costs of measuring the missing outcomes. (DOC 56 kb)

Abbreviations

CI: Confidence interval; COMET: Core Outcome Measures in Effectiveness Trials; COS: Core outcome sets; ICTRP: International Clinical Trials Registry Platform; RCT: Randomized controlled trial; SoF: Summary of findings; WHO: World Health Organization

Acknowledgments

We thank Elise Diard for help in designing Fig. 2.

Funding

This study did not receive any specific funding.

Availability of data and materials

Data available upon request by academic researchers.

Authors' contributions

YY was involved in study conception, selection of trials, data extraction, data analysis, interpretation of results, and drafting the manuscript. AD was involved in study conception, selection of trials, data extraction, data analysis, interpretation of results, and drafting the manuscript. IA was involved in initial data management, interpretation of results, and drafting the manuscript. VT, IB, and PC were involved in the interpretation of results and drafting the manuscript. PR was involved in study conception, interpretation of results, and drafting the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Ethics approval was not applicable. This was a research-on-research study of pre-existing published research; therefore, there was no patient recruitment.

Competing interests

The authors declare that they have no competing interests. All authors have completed the International Committee of Medical Journal Editors uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that they received no support

from any organization for the submitted work, they have no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and they have no other relationships or activities that could appear to have influenced the submitted work.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹INSERM, U1153, Hôpital Hôtel-Dieu, 1, place du parvis Notre Dame, 75004 Paris, France. ²Sorbonne Universités, UPMC Paris Univ-06, Paris, France. ³Service des Urgences - Hôpital Saint Antoine, Assistance Publique-Hôpitaux de Paris (APHP), Paris, France. ⁴Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique-Hôpitaux de Paris (APHP), Paris, France. ⁵Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France. ⁶Cochrane France, Paris, France. ⁷Department of Epidemiology, Columbia University, Mailman School of Public Health, New York, USA.

Received: 30 October 2017 Accepted: 22 May 2018

Published online: 11 June 2018

References

- Tugwell P, Boers M. OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: introduction. *J Rheumatol*. 1993;20(3):528–30.
- Frank L, Basch E, Selby JV. Patient-centered outcomes research I: the PCORI perspective on patient-centered outcomes research. *Jama*. 2014;312(15):1513–4.
- Koroshetz W. A core set of trial outcomes for every medical discipline? *BMJ*. 2015;350:h85.
- Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy*. 2012;17(1):1–2.
- Tunis SR, Clarke M, Gorst SL, Gargon E, Blazeby JM, Altman DG, Williamson PR. Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*. 2016;5(2):193–205.
- Murad MH, Shah ND, Van Houten HK, Ziegenfuss JY, Deming JR, Beebe TJ, Smith SA, Guyatt GH, Montori VM. Individuals with diabetes preferred that future trials use patient-important outcomes and provide pragmatic inferences. *J Clin Epidemiol*. 2011;64(7):743–8.
- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>.
- Ioannidis JPA. Why most clinical research is not useful. *PLoS Med*. 2016;13(6):e1002049.
- Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86–9.
- Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, Chalmers I. Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. 2014;383(9912):176–85.
- Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, Howells DW, Ioannidis JP, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383(9912):156–65.
- Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotsche PC, Krumholz HM, Ghersi D, van der Worp HB. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383(9913):257–66.
- Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267–76.
- Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383(9912):166–75.
- Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101–4.
- Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PMM, Korevaar DA, Graham ID, Ravaud P, Boutron I. Increasing value and reducing waste in biomedical research: who's listening? *Lancet*. 2016;387(10027):1573–86.
- Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ*. 2015;350:h809.
- Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. *J Clin Epidemiol*. 2014;68(3):237–45.
- Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, Swiglo BA, Isley WL, Guyatt GH, Montori VM. Patient-important outcomes in registered diabetes trials. *JAMA*. 2008;299(21):2543–9.
- Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. *BMJ*. 2010;341:c5707.
- Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern Med*. 2013;173(8):611–2.
- Naci H, Ioannidis JP. How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? *Annu Rev Pharmacol Toxicol*. 2015;55:169–89.
- Ocana A, Tannock IF. When are "positive" clinical trials in oncology truly positive? *J Natl Cancer Inst*. 2010;103(1):16–20.
- Glud C, Krogsgaard K. Would you trust a surrogate respondent? *Lancet*. 1997;349(9053):665–6.
- The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med*. 1989;321(6):406–12.
- Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, Taylor RS. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ*. 2013;346:f457.
- Moynihan R. Surrogates under scrutiny: fallible correlations, fatal consequences. *BMJ*. 2011;343:d5160.
- Walter SD, Sun X, Heels-Ansdell D, Guyatt G. Treatment effects on patient-important outcomes can be small, even with large effects on surrogate markers. *J Clin Epidemiol*. 2012;65(9):940–5.
- Chan AW, Hrobjartsson A, Haahr MT, Gotsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457–65.
- Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR. Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database Syst Rev*. 2011;1:MR000031.
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 2009;302(9):977–84.
- Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med*. 2005;24(10):1547–61.
- Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365.
- Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC Med*. 2015;13:282.
- Chan AW, Kileza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ*. 2004;171(7):735–40.
- Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753.
- Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. *BMC Med*. 2016;14(1):100.
- Mocellin S, Lens MB, Pasquali S, Pilati P, Chiarion Sileri V. Interferon alpha for the adjuvant treatment of cutaneous melanoma. *Cochrane Database Syst Rev*. 2013;6:CD008955.
- Ameur H, Ravaud P, Fayard F, Riveros C, Dechartres A. Systematic reviews of therapeutic interventions frequently consider patient-important outcomes. *J Clin Epidemiol*. 2017;84:70–77.
- Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA*. 2007;297(5):468–70.
- Wuytack F, Smith V, Clarke M, Williamson P, Gargon E. Towards core outcome set (COS) development: a follow-up descriptive survey of outcomes in Cochrane reviews. *Syst Rev*. 2015;4(1):73.

42. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative [<http://www.comet-initiative.org/>]. Accessed 11 Apr 2017.
43. Methodology Committee of the Patient-Centered Outcomes Research I. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA*. 2012;307(15):1636–40.
44. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *JAMA*. 2012;307(15):1583–4.
45. Clarke M, Williamson P. Core outcome sets and trial registries. *Trials*. 2015;16(1):216.
46. Gargon E, Williamson PR, Altman DG, Blazeby JM, Tunis S, Clarke M. The COMET initiative database: progress and activities update (2015). *Trials*. 2017;18(1):54.
47. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET initiative database: progress and activities update (2014). *Trials*. 2015;16:515.
48. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET initiative database: progress and activities from 2011 to 2013. *Trials*. 2014;15:279.
49. 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.
50. Peters JL, Mengersen KL. Meta-analysis of repeated measures study designs. *J Eval Clin Pract*. 2008;14(5):941–50.
51. Useem J, Brennan A, LaValley M, Vickery M, Ameli O, Reinen N, Gill CJ. Systematic differences between Cochrane and non-Cochrane meta-analyses on the same topic: a matched pair analysis. *PLoS One*. 2015;10(12):e0144980.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Additional file 1: Summary of the inclusion and **exclusion** criteria at the different phases of the study

Methods phase	Inclusion criteria	Exclusion criteria
I. Identification of RCTs excluded from meta-analyses	<u>Cochrane reviews:</u> - Examining the effects of healthcare interventions (pharmaceutical and non-pharmaceutical) - Published between March 2011 and September 2014 - Reporting a Summary of findings table (SoF). <u>RCTs:</u> - Included in at least one meta-analysis of the main comparison - Excluded from at least one meta-analysis of an important outcome (i.e., an outcome reported in the SoF table)	<u>Cochrane reviews:</u> - Withdrawn - Including observational studies - Including only trials published before 2007 <u>RCTs:</u> - Not included in any meta-analyses of the main comparison
II. Evaluation of the reason for missing outcome	RCTs Published in 2010 or later	RCTs - With inaccessible reports (e.g. retracted publications) - Published in languages other than English or French.
III. Evaluation of research waste	RCTs - With at least one outcome classified as “Selective reporting” at phase II. - With at least one outcome classified as “Incomplete reporting” at phase II. - With at least one outcome classified as “Inadequate planning” at phase II.	

Additional file 2: Qualifications and areas of expertise of the experts involved in the final step of the study

The 5 experts involved in the final step of our study were all physicians and researchers, with different medical specialties and various backgrounds.

1. Dr Agnes Dechartres is a public health specialist, has published many meta-epidemiological or methodological studies, and has been involved as a methodologist in several clinical trials.
2. Pr Isabelle Boutron is a rheumatologist, has published many studies pertaining to research on research and has been involved as a methodologist in several clinical trials.
3. Dr Perrine Crequit is a pneumologist and has published several studies in the field of pneumology, oncology and systematic reviews. She participated as a local investigator of several clinical trials.
4. Pr Philippe Ravaud is a rheumatologist and has published many clinical trials and studies pertaining to research on research. He has been involved as a PI and a methodologist in numerous clinical trials.
5. Dr Viet-Thi Tran is a general practitioner and has worked extensively on topics such as the burden of treatment or treatment adherence in multimorbid patients. He was involved in several clinical trials, including studies led in challenging areas such as Sub-Saharan Africa.

Additional file 3: Characteristics of the 820 Cochrane systematic reviews and the analysed subset of 290 reviews:

Cochrane review groups	All eligible reviews N=820	Sample N=290
	No. (%)	No. (%)
Airways	67 (8)	21 (7)
Menstrual disorders and subfertility	51 (7)	20 (7)
Anaesthesia	46 (6)	18 (6)
Hepato-biliary	43 (5)	12 (4)
Oral health	37 (5)	15 (5)
Infectious diseases	29 (4)	10 (3)
Musculoskeletal	28 (3)	11 (4)
Schizophrenia	28 (3)	15 (5)
Endocrinology	25 (3)	7 (2)
Other	466 (57)	156 (54)

Additional file 4: Experts' opinion of the feasibility and costs of measuring the missing outcomes

Measuring the outcome: trialist perspective	n=282
	No. (%)
Easy	227 (81)
Moderately easy	50 (17)
Difficult	3 (1)
Impossible	0 (0)
Already available	2 (1)
Measuring the outcome: patient perspective	n=282
	No. (%)
Easy	229 (82)
Moderately easy	48 (16)
Difficult	3 (1)
Impossible	0 (0)
Already available	2 (1)
Measuring the outcome: cost evaluation	n=282
	No. (%)
No cost	221 (79)
Minor costs	48 (16)
Moderate costs	11 (4)
Major costs	2 (1)
Outcome importance	n=282
	No. (%)
Important	221 (78)
Not Important	59 (21)
Already available	2 (1)
How easy would it be to measure it as compared with the other measured or planned outcomes?	n=282
	No. (%)
Easier	62 (22)
As easy	165 (59)
More difficult	53 (18)
Already available	2 (1)

4. Discussion et perspectives

4. 1. Discussion

4.1.1. Principaux résultats

Dans le cadre de ce travail de thèse, nous avons évalué le gaspillage de la recherche lié à une planification inappropriée de celle-ci. Nous avons successivement analysé 1) le gaspillage de la recherche dû à l'emploi de méthodes inadéquates dans des essais cliniques puis avons estimé dans quelle mesure ce gaspillage pouvait être évité de manière simple et peu coûteuse ; 2) le gaspillage, évitable, de la recherche lié au choix et à la présentation des critères de jugement dans les essais cliniques, après avoir identifié les critères de jugement les plus importants à mesurer dans le cadre de l'évaluation d'une thérapeutique. Nos résultats soulignent l'importance du problème du gaspillage de la recherche biomédicale tout en suggérant qu'une part de celui-ci pourrait être évité.

Le premier travail a permis de souligner qu'une partie importante des essais cliniques inclus dans des revues Cochrane récentes pouvaient être considérés comme à risque de biais élevé (43 %) ou à risque de biais incertain (41 %), et ainsi être considérés comme gaspillés. Mais que ce gaspillage était partiellement évitable en appliquant des ajustements méthodologiques simples et peu coûteux. De tels ajustements pourraient réduire le risque de biais dans la moitié des essais et l'abroger dans 12 % des essais (IC95% 7 % -18 %). Dans une étude de simulation

corrigeant les rapports incomplets, ce gaspillage évitable représentait jusqu'à 42 % des essais (IC 95 % 36 % à 49 %).

Dans la seconde partie de cette thèse, nous avons abordé la question du gaspillage de la recherche lié au choix des critères de jugement dans les essais. Nous avons évalué les critères de jugement de revues Cochrane d'un grand nombre de spécialités médicales et avons déterminé que la grande majorité des critères de jugement des SoF pouvaient être considérés comme des critères de jugement importants pour les patients (77 %, IC 95 % : 75-79). Et que la quasi-totalité des revues incluait au moins un critère de jugement important pour les patients dans le SoF (96 %, IC 95 % : 93-98). Puis, nous avons estimé la part de gaspillage évitable de la recherche liée aux choix et à la présentation des critères de jugement dans les essais cliniques en prenant comme référence les critères rapportés dans les SoF par les auteurs des revues Cochrane. Nous avons ainsi pu déterminer que de nombreux essais n'ont pas été inclus dans l'ensemble des méta-analyses des critères de jugement importants (78 %, IC 95 % : 76-80), et que chaque ECR n'a contribué qu'à la moitié des méta-analyses, en moyenne. Dans près de 80 % des essais, une planification inadéquate expliquait le manque des CJI, alors que dans 40 % et 15 % des ECRs les rapports incomplets et une présentation sélective rendaient l'inclusion dans la méta-analyse impossible. Ce gaspillage de la recherche, lié aux critères de jugement, aurait pu être partiellement évité pour 63 % des essais et totalement pour 30 %.

4.1.2. Implications

Planification des essais cliniques

L'équipoise clinique est un prérequis nécessaire à la conduite des essais cliniques. Dans le cas de l'évaluation thérapeutique, elle existe quand il n'y a pas de consensus sur les mérites comparés des interventions testées. Un essai ne sera éthiquement acceptable que si au début de celui-ci il existe une equipoise clinique quant aux mérites des différents traitements et que l'essai est planifié de manière à ce qu'il soit raisonnable d'en attendre une modification de celle-ci (260). Un essai biaisé, du fait d'une planification inadéquate, représente une « opportunité manquée ». L'équipoise clinique demeurera inchangée après celui-ci.

La recherche clinique est un processus séquentiel, chaque nouvel essai va contribuer à l'ensemble des éléments de preuve existants. Les évolutions thérapeutiques en médecine proviennent d'un consensus bâti de manière progressive dans les communautés de chercheurs et soignants. Ainsi, si un essai échoue à cause d'une planification inadéquate, il représente une opportunité manquée de participer à ce consensus progressif et sera donc source de gaspillage de la recherche, et de ressources, au niveau de la méta-analyse parce qu'il ne contribuera pas à la preuve globale. Une opportunité manquée au niveau de l'essai individuel, par son incapacité à remettre en cause l'équipoise clinique, entraînera un gaspillage de la recherche au niveau collectif.

Nos travaux suggèrent qu'une part du gaspillage de la recherche serait évitable. Au cours de la phase de planification, des choix méthodologiques tels que le mode de génération de la séquence de randomisation, l'assignation secrète, le choix des critères de jugement importants

ou encore l'utilisation de certaines méthodes statistiques pourrait simplement, mais partiellement, limiter celui-ci.

Des ajustements simples et peu coûteux des méthodes permettraient de réduire le risque de biais de près de 50 % des essais. Une des explications possibles à la persistance de ces biais serait l'absence d'appui méthodologique au cours de la phase de planification des ECRs ou encore l'ignorance des chercheurs des recommandations en termes de méthodes de recherche (77, 230). Mais dans plus de 21 % des essais que nous avons évalués, les problèmes méthodologiques portaient sur des sujets aussi fondamentaux, et simples, que la génération de la séquence de randomisation ou l'allocation secrète. Or les bonnes pratiques dans ce domaine ont souvent été décrites et confirmées par les études métaépidémiologiques(127, 129). La profusion de ressources (24, 25, 261-267), mais aussi la nature du plan expérimental majoritairement utilisé (essai de supériorité, en groupes parallèles, avec un ratio d'allocation de 1:1)(233), devrait aider les chercheurs planifiant ces essais à le faire de manière appropriée, en limitant les sources de biais éventuels. Mais nos résultats semblent indiquer le contraire.

On se doit alors d'imaginer différentes solutions parce que celles disponibles ne suffisent pas. Certaines semblent peu réalistes au premier abord, comme par exemple donner accès à tous les chercheurs qui le désireraient à un support méthodologique. Mais il est fort probable que les développements technologiques récents, en particulier dans le domaine de l'intelligence artificielle et du Natural Language Processing (NLP), pourraient rendre ces solutions, impossibles il y a quelques années, accessibles aujourd'hui. Le NLP, ou traitement automatisé

du langage naturel, est une branche majeure du Machine Learning et donc de l'intelligence artificielle. Il désigne la capacité d'un programme à comprendre le langage humain. Des exemples concrets de NLP nous entourent déjà tels que les assistants personnels développés par les GAFA (acronyme désignant Google, Apple, Facebook et Amazon et Microsoft) ou plus simplement les algorithmes de gestion des courriels dans Google mail. Il est possible d'imaginer que ces outils puissent analyser, et détecter de manière automatisée les principaux traits d'un essai clinique en se basant sur son protocole. La rédaction de ce dernier pouvant être standardisée en se basant sur les recommandations SPIRIT (Standard Protocol Items : Recommendations for Interventional Trials), parues en 2013, et qui définissent ce que doit contenir au minimum un protocole d'essai clinique(265-267). Ces dernières préconisent une description complète de l'essai envisagé, à l'aide d'exemples concrets, sans pour autant suggérer la meilleure manière de le concevoir ou de le mener. Selon les éléments mis en évidence, la plateforme serait alors susceptible de proposer des amendements au protocole, pour limiter le risque de biais en se basant sur les recommandations citées précédemment (24, 25, 261-267), et résultats des études métaépidémiologiques (129, 130, 134, 135, 159, 245, 268).

Une solution avoisinante semble proposée par trials.ai. Il s'agit d'un start-up californienne qui affirme utiliser l'intelligence artificielle et le NLP pour aider les chercheurs à gérer plus efficacement la conduite des essais cliniques. Les algorithmes de leur plate-forme, appelée « Study Optimizer », auraient été développés à partir « des milliards de points de données d'essais cliniques antérieurs, de données de la littérature et de sources du monde réel ». Ces algorithmes doivent ainsi identifier les obstacles à la réussite des ECRs et fournir des recommandations pour leur optimisation. Ces obstacles semblent être essentiellement d'ordre organisationnel (optimiser le recrutement, minimiser les perdus de vue, mise à disposition de Case Report Form électronique...). Peu d'informations sont disponibles sur cette plateforme,

qui affirme par exemple avoir aidé le « Moores Cancer Center » de l'Université de Californie à San Diego à réduire la durée des essais de 33 % et réduire les erreurs de données de 20 % (269). Ils semblent s'adresser essentiellement aux gestionnaires de la recherche au sein de l'industrie afin d'optimiser les coûts des essais cliniques, mais la partie méthodologique semble peu ou pas abordée (269). Il est possible d'envisager des solutions similaires, qui seraient capables de répondre et d'aider les chercheurs menant des essais cliniques académiques et/ou sans soutien méthodologique. Une telle offre permettrait également d'aider les centres experts, tels que les « unités de recherche clinique » ou les « centres de recherche et d'investigation », en automatisant une partie de l'analyse des protocoles que ceux-ci reçoivent, pour consacrer leur expertise aux cas les plus complexes.

De nombreux essais que nous avons évalués présentaient des problèmes méthodologiques liés aux statistiques. Ils touchaient en particulier les méthodes d'analyses des essais cliniques et la gestion des données manquantes. Dans un sondage auprès d'éditeurs de journaux biomédicaux, Fernandes-Taylor *et al* rapportent que les principaux problèmes statistiques retrouvés dans les articles soumis concernaient l'utilisation de plans statistiques inappropriés ou une mauvaise gestion des données manquantes (270). D'un côté, la littérature en biostatistique est riche d'idées nouvelles et innovantes (28, 271-275). De nombreux travaux par exemple décrivent les méthodes permettant de limiter le biais d'attrition par une gestion appropriée des données manquantes (273-275). D'un autre côté, la multiplication des logiciels de statistique, à but commercial ou sous forme de logiciel libre, a démocratisé ces outils (276). Calculer le nombre de sujets nécessaires pour un essai ou encore faire les principaux tests est facilement réalisable à l'aide d'un site internet gratuit, se basant le logiciel de statistique R (277). Mais l'expertise

nécessaire à l'établissement d'un plan d'analyse statistique adéquat et à la bonne utilisation de ces outils reste importante, et partielle, chez les chercheurs sans formation spécifique (230, 278). La majorité des analyses statistiques retrouvées dans la littérature restent les techniques statistiques de base ou des statistiques descriptives (279-281). Et celles-ci semblent insuffisamment rapportées et/ou maîtrisées (279-281). Ces éléments mettent en évidence la nécessité pour les chercheurs et les professionnels de santé de bien comprendre les statistiques de base avant d'utiliser les outils à leur disposition. Et de se rapprocher de statisticiens lors de l'utilisation de techniques plus avancées, telle que la gestion des données manquantes, l'utilisation des scores de propension ou les méta-analyses.

Nos travaux ont également décrit qu'une partie du gaspillage de la recherche était liée à la planification des critères de jugement et à l'incapacité de nombreux essais à mesurer l'ensemble des critères de jugement importants. Nos résultats retrouvent que la majorité des critères de jugement rapportés dans les SoF tables des revues Cochrane pourrait être considérés comme des CJI. Ce résultat corrobore des travaux de recherche récemment publiés et évaluant les critères de jugement des revues systématiques (250-252). Sur un échantillon de 2015 revues, toutes les revues Cochrane incluaient au moins un critère de jugement important pour les patients, tandis que 71 % des revues publiées en 2010 avec un CJIP comme critère de jugement principal (250-252). Cependant il existe une disparité importante entre les critères choisis dans les ECRs et ceux des revues systématiques. Moins du quart des essais sur le diabète et des essais de cardiologie ont utilisé un CJIP comme critère de jugement principal (149, 150), même si les patients indiquent une forte préférence pour les essais évaluant l'effet des interventions thérapeutiques sur des CJIP (244). Dans nos travaux seulement 22 % des revues avaient au

moins un CJIP avec un niveau de preuve élevé, ce qui peut indiquer un manque d'éléments de preuves pour ces critères de jugement. Une enquête sur les revues Cochrane publiées entre 2007 et 2011 a révélé que 37 % des critères de jugement prédéfinis dans les méthodes des revues n'étaient pas rapportés dans les résultats, poussant les auteurs à supposer qu'une partie de ces essais n'avaient probablement pas mesuré ou rapporté ces critères de jugement (169, 250). Ces hypothèses sont confirmées par nos travaux avec uniquement 58 % des ECRs de la comparaison principale ayant contribué aux méta-analyses de critères de jugement du SoF Table. Les essais étaient exclus des méta-analyses par défaut de planification dans près de trente pour cent des situations (29 %, n = 282) ou pour des résultats présentés de manières sélective ou incomplète dans 25 % des cas.

L'identification des critères de jugement importants, ou les plus importants, à mesurer dans une situation est difficile et il est évident que sans un consensus sur ceux-ci cette hétérogénéité persistera (256). Des initiatives internationales se mettent en place pour tenter d'y remédier. Aux États-Unis, l'institut PCORI (Patient-Centered Outcomes Research Institute) finance et promeut la recherche comparative (comparative effectiveness research CER) qui vise à améliorer la qualité des soins tout en mesurant des critères de jugement importants et significatifs pour les patients (146-148). L'initiative COMET (Core Outcome Measures in Effectiveness Trials) encourage le développement et l'utilisation des COS, qui pourraient aider les chercheurs dans leur choix de critères de jugement à la phase de planification. L'ICHOM (International Consortium for Health Outcome Measurement) travaille également sur l'établissement de « standard » de critères de jugement (282). Ces initiatives doivent permettre d'améliorer la pertinence, diminuer l'hétérogénéité des critères de jugement et ainsi participer à lutter contre le gaspillage de la recherche (242, 283, 284). Mais cette ambition se heurte à deux écueils majeurs : 1) le nombre de COS développés ou en cours de développement reste

réduit (285-287) 2) l'outil proposé par la COMET est peu utilisable en l'état. Il semble donc nécessaire que le site de la COMET initiative évolue. Pour que les COS puissent impacter la recherche clinique ils doivent être accessibles et utilisables, or une recherche sur cette plateforme donne un résultat confus et très parasité par des protocoles de COS en cours, protocole de revues systématiques, commentaires et avis divers. Une présentation rudimentaire, tabulaire, de ceux déjà disponibles serait une possible solution, simple et réalisable. Mais celle-ci restera limitée par le nombre de COS disponibles, leur qualité non évaluée et par la complexité du processus du développement de ceux-ci (288).

Rapports incomplets et présentation sélective des résultats

Pour pouvoir correctement interpréter les résultats d'un essai clinique, ils doivent être présentés de manière claire, exhaustive et transparente. Une grande partie des ECRs inclus dans des revues Cochrane récentes avaient un risque de biais incertain (41 %) ou ne pouvaient pas être inclus dans des méta-analyses du fait de critères de jugement rapportés de manière sélective ou incomplète (30 %). Bien que plusieurs publications aient souligné que la qualité du rapport d'un essai ne reflète pas nécessairement la façon dont il a été mené (un mauvais rapport ne signifie pas forcément que l'ECR était biaisé), un rapport complet est préalable absolument nécessaire à une évaluation critique des essais (235, 289).

Le moyen le plus simple et le moins coûteux de réduire le gaspillage de la recherche reste de rapporter correctement les détails méthodologiques et les résultats d'un essai clinique. Ces

détails sont en général disponibles auprès des auteurs correspondants et devraient être mis à disposition dans les rapports d'essais ou à défaut dans des appendices accessibles (215, 218-221). Les résultats et les critères de jugement étant de leur côté plus précisément rapportés sur ClinicalTrials.gov que dans les articles (259), il est possible d'imaginer que le problème de la présentation incomplète de ceux-ci serait également simple et peu coûteux à corriger. Depuis plusieurs années, d'importants efforts ont été faits pour aider les auteurs à améliorer la qualité et l'exhaustivité des articles de recherche. Dechartres *et al*, après avoir analysé 20920 ERC, extraits de 2001 revues systématiques et publiés dans 3136 journaux, rapportent par exemple que la qualité des rapports des essais semblait s'améliorer avec le temps (290). Mais cette amélioration n'intéresserait pas tous les éléments des rapports, la génération de la séquence de randomisation ou l'allocation secrète étant les domaines ayant le plus bénéficié de cette transparence accrue. De nombreuses recommandations ont été développées et sont activement promues par le réseau EQUATOR (Enhancing the QUALity and Transparency Of health Research) (25, 236, 237). Bien que leur utilisation demeure suboptimale (238-240), il existe des outils pour parfaire les rapports d'essais cliniques et augmenter l'adhésion aux recommandations (240, 259, 291). La participation active des éditeurs à l'implémentation de celles-ci permettrait par exemple d'améliorer la qualité des abstracts (240, 292). Un outil d'aide à la rédaction, tel que COBWEB (Consort-based WEB tool), a aussi permis de considérablement améliorer la description de la section méthodes d'ECR, bien que son effet sur l'exhaustivité des rapports demeure encore incertain (293).

4.1.3. Limites

Impact de la méthode d'évaluation du risque de biais

Dans la première partie de cette thèse, nous avons utilisé une approche originale en proposant de limiter, lors de la phase de planification, les éléments méthodologiques susceptibles d'entraîner un risque de biais élevé. Mais ce travail reposait en très grande partie sur l'évaluation du risque de biais dans les ECRs, à l'aide du RoB tool, par les auteurs de revues systématiques Cochrane, ou par nous pour un échantillon aléatoire de 200 essais à risque de biais élevé. Or, bien que le RoB tool soit reconnu et recommandé par la communauté internationale, il comporte néanmoins d'importantes limites. Plusieurs travaux ont décrit que le RoB tool souffrait d'une reproductibilité globalement faible (294-297). Ce qui semble à nouveau se confirmer selon nos résultats, avec un désaccord entre les auteurs des revues et notre évaluation pour 29 % des essais. Cette mauvaise reproductibilité touchait en particulier les domaines nécessitant un jugement, avec une part de subjectivité de la part de l'évaluateur, tel que l'aveugle ou la gestion des données manquantes (294-297). Afin d'en limiter les conséquences sur nos conclusions, nous avons réestimé le risque de biais en tenant compte du caractère objectif ou subjectif des critères de jugement.

Afin d'améliorer sa reproductibilité, il a été suggéré de procéder à l'évaluation du risque de biais de manière indépendante, en double, et avec consensus. Mais malgré cette méthode, la reproductibilité restait faible, au sein et entre les binômes d'évaluateurs (295). De nombreuses raisons ont été avancées pour expliquer cette mauvaise reproductibilité, comme le manque de définitions précises concernant certains domaines (p. ex. pas de limite chiffrée à la quantité de données manquantes acceptables) ou l'insuffisance de formation des auteurs de revues à

l'utilisation de cet outil. Ainsi, le RoB tool était appelé à évoluer (298). En septembre 2018, la Collaboration Cochrane a mis à disposition une version préliminaire du futur RoB Tool, baptisée « RoB 2,0 tool (revised tool for Risk of Bias in randomized trials) » (299). Celui-ci, à l'aide d'algorithmes et de questions pivots (« signalling questions ») servant à guider les évaluateurs, permettrait possiblement de réduire la variabilité et d'améliorer la reproductibilité de l'évaluation du risque de biais. Le Rob 2,0 tool a également été décliné en plusieurs versions pour s'adapter aux essais en grappes (cluster trials), aux essais croisés (cross over trials). Il existe enfin une version spécifique du RoB tool pour les études interventionnels non randomisés (ROBINS-I tool [Risk Of Bias in Non-randomized Studies — of Interventions]).

Impact de la source des données : revues systématiques Cochrane

Dans l'ensemble des travaux de cette thèse, nous nous sommes basés sur des essais cliniques inclus dans des revues systématiques Cochrane. De par ce choix, il est probable que nous ayons sous-estimé le gaspillage de la recherche dû à une planification inadéquate des essais. En effet, les revues systématiques Cochrane ont été décrites comme étant de meilleure qualité que les revues non Cochrane, tout en incluant moins d'essais cliniques (21, 300-302). Il est donc probable que les auteurs de revues aient exclus des essais ne satisfaisant pas à certains standards méthodologiques (par exemple essais dits quasi randomisés). De ce fait, il est probable que notre proportion d'essais à risque de biais élevé ait été une sous-estimation de la proportion réelle d'essais à risque de biais élevé.

Ensuite, dans notre second travail, nous nous sommes concentrés sur un échantillon d'essais inclus dans les revues Cochrane, mais exclus d'au moins une méta-analyse. Nous n'avons donc pas examiné les essais exclus de la revue systématique pour des raisons de critères de jugement non prévus, non ou mal rapportés. Or, bien que les recommandations du Cochrane Handbook visant à déterminer l'éligibilité des essais soient très claires sur ce point (les études ne doivent pas être exclues simplement parce qu'elles ne fournissent pas de données « utilisables ») (29), certains essais peuvent avoir été exclus pour cette raison. De ce fait, il est probable que nous sous-estimons le gaspillage lié aux critères de jugement. Il est enfin possible de supposer que cette proportion aurait été encore plus importante dans des revues non Cochrane.

Impact de l'évaluation par les experts

Pour l'ensemble de nos travaux, nous avons fait appel à l'évaluation et au jugement, d'un groupe d'experts. Ceux-ci sont intervenus pour valider les ajustements méthodologiques proposés, déterminer leur faisabilité et estimer leur coût éventuel dans le premier travail. Puis un autre groupe d'experts a été amené à évaluer la faisabilité de mesure des critères de jugement importants non prévus (planification inadéquate), ainsi que leur coût éventuel. Certains experts ont participé aux deux travaux.

Les experts impliqués sont reconnus dans le domaine des méthodes des essais cliniques, des essais cliniques ou encore de l'évaluation du risque de biais. Dans les deux cas, les réunions des experts ont été modérées par l'un d'entre nous afin de s'assurer que tous les experts aient d'abord donné leur avis, puis débattu ensemble, avant de prendre une décision par consensus,

ceci afin d'éviter que l'opinion de l'une des personnalités devienne prépondérante sur celle des autres.

Pour s'assurer qu'ils n'ont pas été influencés lors de leur évaluation, ils n'ont pas été informés de la proportion de problèmes méthodologiques dans le premier échantillon ou encore les proportions de critères manquants dans le second. Mais, malgré ces mesures, cette évaluation reste un exercice difficile et expose à une subjectivité certaine.

Impact du choix des critères de jugement dit importants

Réunir des experts de l'ensemble des spécialités explorés par les essais cliniques inclus dans les travaux de cette thèse aurait été probablement impossible. Nous avons donc considéré que les critères de jugement du SoF table était les critères de jugement les plus importants à mesurer dans une situation donnée, pour une intervention donnée et dans une pathologie donnée. Nous avons utilisé les critères du SoF Table comme des critères de substitutions (« surrogate outcomes »). Les critères du SoF table sont choisis par les auteurs des revues, et pourraient ne pas être nécessairement les critères les plus importants dans le domaine, ces derniers n'étant pas nécessairement des experts en contenu, mais des experts en médecine basée sur les preuves ou en méthode de revues systématiques (303).

4.2. Conclusion

Ce travail de thèse suggère que le gaspillage de la recherche pourrait être en partie évité au cours de la phase de planification des essais cliniques par des ajustements méthodologiques simples et peu coûteux et par le choix des critères de jugement à mesurer. Mais, comme nous l'avons mis en évidence, ces ajustements ne sont pas suffisants. Une grande partie des rapports d'essais et des critères de jugement souffraient d'un manque de transparence. Or rapporter un essai et l'ensemble de ses critères de jugement de manière claire et transparente est l'ajustement le plus simple et le moins cher envisageable.

4. 3. Perspectives

Nous allons prolonger nos travaux selon plusieurs axes. D'une part nous allons poursuivre notre étude du gaspillage de la recherche au sein des essais cliniques et d'autre part nous nous intéresserons au gaspillage de la recherche au sein des revues systématiques.

Gaspillage de la recherche dans les essais cliniques : priorisation et critères de jugement

Nous prévoyons dans un premier temps de nous intéresser au gaspillage de la recherche lié à la définition des priorités de recherche. Comme nous l'avons précédemment souligné, une part du gaspillage de la recherche serait dû à des projets de recherche ne répondant pas aux besoins des utilisateurs de celle-ci. Des initiatives telles que la James Lind Alliance, à l'aide de « Priority Setting Partnership » (PSP), cherchent à identifier et hiérarchiser les incertitudes thérapeutiques ou les questions sans réponse pour des pathologies et des domaines de soins spécifiques. Une initiative de priorisation de la recherche a été menée par la JLA, en collaboration avec le « Royal College of Emergency Physicians », dans le champ de la médecine d'urgence en 2017 (304, 305). Cette initiative a donné lieu à la publication d'une liste de 72 questions de recherche considérées prioritaires. Nous allons chercher de manière systématique les registres d'essais cliniques (ClinicalTrials.gov, International Clinical Trials Registry Platform, EU Clinical Trials Register etc.) pour identifier un échantillon d'essais enregistrés dans le champ de la médecine d'urgence. De manière concomitante, nous comparerons ces questions de recherche aux revues systématiques Cochrane publiées dans le champ de la médecine d'urgence. Ce champ de la médecine étant très large, nous nous baserons sur les travaux du PEC field (« Pre-hospital and Emergency Care field ») de la Cochrane, donc l'une des tâches consiste à identifier l'ensemble

des revues systématiques intéressant les champs de la médecine d'urgence intra et pré hospitalière. Ensuite nous évaluerons si les essais enregistrés rentrent dans le cadre de l'une des priorités établies par la JLA et le RCEM ; ou cherchent à répondre à des incertitudes mises en avant par les conclusions des revues Cochrane identifiées par le PEC Field.

Dans un second temps nous nous intéresserons à nouveau au gaspillage de la recherche lié aux choix des critères de jugement. Celui-ci pourrait être réduit par l'utilisation d'un ensemble de critères de jugement commun pour les essais portant sur certaines pathologies, dans certains contextes. L'initiative COMET propose une méthodologie pour développer des COS (Core Outcome Sets), mais celle-ci est longue et complexe. Nous faisons l'hypothèse que les critères de jugement des SoF tables pourraient servir de COS de substitution dans les domaines où ceux-ci sont manquants. Pour vérifier celle-ci, nous nous baserons sur le site <http://www.comet-initiative.org/>, nous y identifierons les champs communs entre la base de données de SoF tables de revues Cochrane que nous avons précédemment constituées (253, 254) et les COS disponibles. Puis nous comparerons celle-ci aux critères de jugement issus des SoF tables.

Gaspillage de la recherche dans les revues systématiques

D'autre part, certains auteurs ont récemment suggéré que les revues systématiques pouvaient également participer d'une certaine manière au gaspillage de la recherche (306-308). De la même manière que des chercheurs ont mis en doute l'intérêt des recherches systématiques chronophages et coûteuses pour identifier les données issues de la littérature grise ou de bases

de données non anglophones(120), d'autres font valoir que les essais à faible effectif sont souvent mal entrepris, mal rapportés et avec un mauvais ratio effet sur biais. Ils représenteraient un abus de l'altruisme des participants et pourraient même participer d'une certaine manière au gaspillage des ressources de la recherche en mobilisant des légions d'évaluateurs pour les examiner, dans l'espoir d'en extraire des peu probables informations (306).

Ces petits essais présentent souvent des effets traitements importants qui rapidement diminuent, voire disparaissent, lorsque des essais de qualité et de grande taille sont réalisés (306, 307). Les efforts déployés par les auteurs de revues pour identifier de manière systématique tous les essais auraient donné une audience inespérée à ces essais de faible qualité méthodologique, monocentriques et peu fiables. La plupart des méta-analyses sont dominées par de tels essais, avec un nombre médian d'essais dans les revues Cochrane de 7, et un nombre médian de patients de l'ordre de 80 (135). L'inclusion de ces essais dans les méta-analyses pourrait conduire à une surestimation des effets traitements (134). Un principe général incessible (rechercher et inclure systématiquement tous les essais) est attrayant, mais devient problématique lorsque, pour des raisons de simplicité, il ne prend pas en compte des aspects importants de la complexité de la nature des preuves disponibles. La conséquence attendue de la volonté d'identifier et d'inclure tous les essais est que certaines études se réclamant comme essais randomisés, ne le soient pas du fait d'une manque de compréhension de ce qu'est la randomisation et la randomisation appropriée par exemple (168, 307). Les méta-analyses des variables de base dans les revues systématiques d'essais randomisés révèlent de surprenants déséquilibres, suggérant une randomisation pour le mieux, inadéquate ou encore absente (309).

Considérer, sans équivoque, tous les essais, quel que soit leur taille ou le nombre d'évènements observés, pourrait involontairement tromper le lecteur de méta-analyse. Il y a 20 ans, Moore *et al* affirmaient que « size is everything » (la taille est tout) : selon eux les estimations crédibles de l'efficacité clinique d'une intervention ne devraient provenir que d'essais de grande envergure (310). Mais la détermination de ce que signifie « grande envergure » n'a pas évolué depuis. Afin d'identifier des pistes pour convenablement prendre en compte les essais à faible effectif dans les revues systématiques, nous évaluerons différentes méthodes d'analyses (méta-analyse limite, analyse limitée aux essais avec le plus grand nombre d'évènements observés, analyse limitée aux essais avec le plus de participants, méta-analyse bayésienne...) en nous basant sur les données individuelles des patients de méga-trials récents (CRASH, CRASH2, WOMEN), ou en cours de finalisation (CRASH3, HALT-IT), afin de simuler des essais de plus ou moins grande taille. Les données de ces essais cliniques multicentriques nous serviront à simuler le comportement des méta-analyses en fonction des méthodes d'analyses utilisées et des critères d'inclusions/exclusions basés sur le nombre de participants. Nous chercherons ainsi à identifier la méthode qui permette d'obtenir l'estimation la plus fiable et la moins hétérogène possible, et ainsi de déterminer une éventuelle taille critique d'inclusion pour les essais au sein des méta-analyses.

Références

1. Nathan T, Stengers I. Médecins et Sorciers: Empêcheurs de penser rond; 2004. 192 p.
2. Morabia A. Claude Bernard was a 19th century proponent of medicine based on evidence. *J Clin Epidemiol*. 2006;59(11):1150-4.
3. Bernard C. Introduction à l'étude de la médecine expérimentale. Paris (1984): Flammarion; 1865.
4. Evidence-Based Medicine Working G. Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama*. 1992;268(17):2420-5.
5. Evidence-based health care: a new approach to teaching the practice of health care. Evidence-Based Medicine Working Group. *J Dent Educ*. 1994;58(8):648-53.
6. 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):71-2.
7. Montori VM, Brito JP, Murad MH. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. *Jama*. 2013;310(23):2503-4.
8. Vlay SC. How the university cardiologist treats ventricular premature beats: a nationwide survey of 65 University Medical Centers. *Am Heart J*. 1985;110(4):904-12.
9. Turner A. Evaluating the UK House of Commons Science and Technology Committee's position on the implausible effectiveness of homeopathic treatments. *Theor Med Bioeth*. 2017;38(4):335-52.
10. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. 2017;390(10092):415-23.
11. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J*. 1979;121(9):1193-254.
12. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1989;95(2 Suppl):2S-4S.
13. Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J*. 1904;3:1243-6.
14. Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. In: *Medicines for the year 2000* London: Office of Health Economics. 1979:1-11.
15. Allen C, Richmond K. The Cochrane Collaboration: international activity within Cochrane Review Groups in the first decade of the twenty-first century. *J Evid Based Med*. 2011;4(1):2-7.
16. Scholten RJPM, Clarke M, Hetherington J. The Cochrane Collaboration. *European Journal Of Clinical Nutrition*. 2005;59:S147.
17. Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*. 1993;703:156-63; discussion 63-5.

18. 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.
19. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evidence Based Medicine*. 2016;21(4):125.
20. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7(9):e1000326.
21. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS Med*. 2016;13(5):e1002028.
22. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-9.
23. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Jama*. 1996;276(8):637-9.
24. Moher D, Schulz KF, Altman D, Group C. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Jama*. 2001;285(15):1987-91.
25. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med*. 2010;7(3):e1000251.
26. Clark S, Horton R. Putting research into context--revisited. *Lancet*. 2010;376(9734):10-1.
27. Roloff V, Higgins JP, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat Med*. 2013;32(1):11-24.
28. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. *Stat Med*. 2007;26(12):2479-500.
29. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. Available from: www.cochrane-handbook.org.
30. Mulrow CD. Rationale for systematic reviews. *BMJ*. 1994;309(6954):597-9.
31. Egger M, Davey Smith G, Altman DG. Systematic reviews in health care: meta-analysis in context. *BMJ Books*. 2001.
32. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011; Available from www.cochrane-handbook.org
33. Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ*. 1997;315(7119):1371-4.
34. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997;315(7121):1533-7.

35. Thomas D, Radji S, Benedetti A. Systematic review of methods for individual patient data meta- analysis with binary outcomes. *BMC Med Res Methodol*. 2014;14:79.
36. Zarychanski R, Abou-Setta AM, Turgeon AF, Houston BL, McIntyre L, Marshall JC, et al. Association of hydroxyethyl starch administration with mortality and acute kidney injury in critically ill patients requiring volume resuscitation: a systematic review and meta-analysis. *Jama*. 2013;309(7):678-88.
37. De Oliveira GS, Jr., Castro-Alves LJ, Chang R, Yaghmour E, McCarthy RJ. Systemic metoclopramide to prevent postoperative nausea and vomiting: a meta-analysis without Fujii's studies. *Br J Anaesth*. 2012;109(5):688-97.
38. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Jama*. 1998;279(4):281-6.
39. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*. 1997;315(7109):640-5.
40. Egger M, Smith GD. Bias in location and selection of studies. *BMJ*. 1998;316(7124):61-6.
41. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252-60.
42. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst*. 1989;81(2):107-15.
43. Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci*. 1993;703:135-46; discussion 46-8.
44. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *Jama*. 1992;267(3):374-8.
45. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev*. 2009(1):MR000006.
46. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001;54(10):1046-55.
47. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ*. 2001;323(7304):101-5.
48. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol*. 2000;53(11):1119-29.
49. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly*. 2008;138(13-14):197-203.
50. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.

51. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337(8746):867-72.
52. Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol*. 1989;36(5):429-32.
53. Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ*. 1997;315(7109):635-40.
54. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)*. 1987;295(6599):654-6.
55. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. *J Clin Epidemiol*. 2002;55(4):407-10.
56. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ*. 1992;305(6844):15-9.
57. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *Jama*. 2004;291(20):2457-65.
58. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753.
59. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365.
60. Cohen D. FDA official: "clinical trial system is broken". *BMJ*. 2013;347:f6980.
61. Spence D. Evidence based medicine is broken. *BMJ : British Medical Journal*. 2014;348.
62. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance G. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
63. Horton R. Offline: What is medicine's 5 sigma? *The Lancet*. 385(9976):1380.
64. Angell M. Drug Companies & Doctors: A Story of Corruption. *The New York Review of Books*. 2009 2009/01/15/.
65. Altman DG. The scandal of poor medical research. *BMJ*. 1994;308(6924):283-4.
66. Ioannidis JPA. Why Most Published Research Findings Are False. *PLoS Med*. 2005;2(8):e124.
67. Ioannidis JPA. Why Most Clinical Research Is Not Useful. *PLOS Medicine*. 2016;13(6):e1002049.
68. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ*. 2012;344:d7292.

69. Woolley KL, Gertel A, Hamilton C, Jacobs A, Snyder G. Poor compliance with reporting research results - we know it's a problem . . . how do we fix it? *Curr Med Res Opin.* 2012;28(11):1857-60.
70. Lehman R, Loder E. Missing clinical trial data. *BMJ.* 2012;344:d8158.
71. Li T, Vedula SS, Scherer R, Dickersin K. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Ann Intern Med.* 2012;156(5):367-77.
72. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet.* 2009;374(9683):86-9.
73. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet.* 2014;383(9912):176-85.
74. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet.* 2014;383(9912):156-65.
75. Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet.* 2014;383(9913):257-66.
76. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* 2014;383(9913):267-76.
77. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* 2014;383(9912):166-75.
78. Kleinert S, Horton R. How should medical science change? *Lancet.* 2014;383(9913):197-8.
79. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet.* 2014;383(9912):101-4.
80. Crowe S, Giles C. Making patient relevant clinical research a reality. *BMJ.* 2016;355:i6627.
81. Jun M, Manns B, Laupacis A, Manns L, Rehal B, Crowe S, et al. Assessing the extent to which current clinical research is consistent with patient priorities: a scoping review using a case study in patients on or nearing dialysis. *Can J Kidney Health Dis.* 2015;2:35.
82. Crowe S, Fenton M, Hall M, Cowan K, Chalmers I. Patients', clinicians' and the research communities' priorities for treatment research: there is an important mismatch. *Research involvement and engagement.* 2015;1:2.
83. Tallon D, Chard J, Dieppe P. Relation between agendas of the research community and the research consumer. *Lancet.* 2000;355(9220):2037-40.
84. Chalmers I. The James Lind Initiative. *J R Soc Med.* 2003;96(12):575-6.
85. Oliver S, Gray J. A bibliography of research reports about patients', clinicians' and researchers' priorities for new research. London: James Lind Alliance. 2006.

86. Corner J, Wright D, Hopkinson J, Gunaratnam Y, McDonald JW, Foster C. The research priorities of patients attending UK cancer treatment centres: findings from a modified nominal group study. *Br J Cancer*. 2007;96(6):875-81.
87. Gross CP, Anderson GF, Powe NR. The relation between funding by the National Institutes of Health and the burden of disease. *N Engl J Med*. 1999;340(24):1881-7.
88. Perel P, Miranda JJ, Ortiz Z, Casas JP. Relation between the global burden of disease and randomized clinical trials conducted in Latin America published in the five leading medical journals. *PLoS ONE*. 2008;3(2):e1696.
89. Stuckler D, King L, Robinson H, McKee M. WHO's budgetary allocations and burden of disease: a comparative analysis. *Lancet*. 2008;372(9649):1563-9.
90. Atal I, Trinquart L, Ravaud P, Porcher R. A mapping of 115,000 randomized trials revealed a mismatch between research effort and health needs in non-high-income regions. *J Clin Epidemiol*. 2018;98:123-32.
91. Rottingen JA, Regmi S, Eide M, Young AJ, Viergever RF, Ardal C, et al. Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory? *Lancet*. 2013;382(9900):1286-307.
92. Emdin CA, Odutayo A, Hsiao AJ, Shakir M, Hopewell S, Rahimi K, et al. Association between randomised trial evidence and global burden of disease: cross sectional study (Epidemiological Study of Randomized Trials--ESORT). *BMJ*. 2015;350:h117.
93. Gillum LA, Gouveia C, Dorsey ER, Pletcher M, Mathers CD, McCulloch CE, et al. NIH disease funding levels and burden of disease. *PLoS ONE*. 2011;6(2):e16837.
94. Liberati A. Need to realign patient-oriented and commercial and academic research. *Lancet*. 2011;378(9805):1777-8.
95. Ahmad N, Boutron I, Dechartres A, Durieux P, Ravaud P. Geographical representativeness of published and ongoing randomized controlled trials. The example of: Tobacco consumption and HIV infection. *PLoS ONE*. 2011;6(2):e16878.
96. Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol*. 2013;13:50.
97. Bath PM, Gray LJ. Systematic reviews as a tool for planning and interpreting trials. *Int J Stroke*. 2009;4(1):23-7.
98. Young C, Horton R. Putting clinical trials into context. *Lancet*. 2005;366(9480):107-8.
99. Ker K, Edwards P, Perel P, Shakur H, Roberts I. Effect of tranexamic acid on surgical bleeding: systematic review and cumulative meta-analysis. *BMJ*. 2012;344:e3054.
100. Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials*. 2005;2(3):218-29; discussion 29-32.

101. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Ann Intern Med.* 2011;154(1):50-5.
102. Cooper NJ, Jones DR, Sutton AJ. The use of systematic reviews when designing studies. *Clin Trials.* 2005;2(3):260-4.
103. Clarke M, Hopewell S. Many reports of randomised trials still don't begin or end with a systematic review of the relevant evidence. *Journal of the Bahrain Medical Society.* 2013;24(3):145-8.
104. Chang SM, Carey TS, Kato EU, Guise JM, Sanders GD. Identifying research needs for improving health care. *Ann Intern Med.* 2012;157(6):439-45.
105. Ioannidis JP, Karassa FB. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. *BMJ.* 2010;341:c4875.
106. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *Jama.* 1992;268(2):240-8.
107. Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *Int J Epidemiol.* 2005;34(4):874-87.
108. Sinclair JC. Meta-analysis of randomized controlled trials of antenatal corticosteroid for the prevention of respiratory distress syndrome: discussion. *Am J Obstet Gynecol.* 1995;173(1):335-44.
109. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med.* 1992;327(4):248-54.
110. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS ONE.* 2014;9(7):e102670.
111. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. *BMJ.* 2013;346:f1135.
112. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *Jama.* 2005;294(2):218-28.
113. Djulbegovic B, Kumar A, Magazín A, Schroen AT, Soares H, Hozo I, et al. Optimism bias leads to inconclusive results-an empirical study. *J Clin Epidemiol.* 2011;64(6):583-93.
114. Bailey KR. Generalizing the results of randomized clinical trials. *Control Clin Trials.* 1994;15(1):15-23.
115. Flather M, Delahunty N, Collinson J. Generalizing results of randomized trials to clinical practice: reliability and cautions. *Clin Trials.* 2006;3(6):508-12.
116. Glasziou P, Guyatt GH, Dans AL, Dans LF, Straus S, Sackett DL. Applying the results of trials and systematic reviews to individual patients. *ACP J Club.* 1998;129(3):A15-6.

117. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005;365(9453):82-93.
118. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med*. 2002;21(11):1513-24.
119. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *Jama*. 2002;287(22):2973-82.
120. Egger M, Juni P, Bartlett C, Hoenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*. 2003;7(1):1-76.
121. Juni P, Tallon D, Egger M. 'Garbage in- garbage out'? Assessment of the quality of controlled trials in meta-analyses published in leading journals. Proceedings of the 3rd symposium on systematic reviews: beyond the basics, St Catherine's college, Oxford Oxford : Centre for Statistics in Medicine. 2000;19.
122. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med*. 2001;135(11):982-9.
123. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609-13.
124. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol*. 2007;36(4):847-57.
125. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama*. 1995;273(5):408-12.
126. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med*. 2007;26(14):2745-58.
127. Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum*. 2009;61(12):1633-41.
128. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med*. 2012;157(6):429-38.
129. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336(7644):601-5.

130. Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ*. 2009;339:b3244.
131. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol*. 2005;34(1):79-87.
132. Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. *Jama*. 2010;304(7):793-4.
133. Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland JM. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. *BMJ*. 2005;330(7499):1057-8.
134. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med*. 2011;155(1):39-51.
135. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ*. 2013;346:f2304.
136. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ*. 2006;332(7547):969-71.
137. Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med*. 2006;99(7):337-41.
138. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
139. Hopewell S, Boutron I, Altman DG, Ravaud P. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *BMJ Open*. 2013;3(8):e003342.
140. Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol*. 2001;54(7):651-4.
141. Tugwell P, Boers M. OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: introduction. *J Rheumatol*. 1993;20(3):528-30.
142. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol*. 1989;42(5):403-8.
143. Guyatt G, Montori V, Devereaux PJ, Schunemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP J Club*. 2004;140(1):A11-2.
144. Richards T, Montori VM, Godlee F, Lapsley P, Paul D. Let the patient revolution begin. *BMJ*. 2013;346:f2614.
145. Montori VM, Gandhi GY, Guyatt GH. Patient-important outcomes in diabetes--time for consensus. *Lancet*. 2007;370(9593):1104-6.

146. Frank L, Basch E, Selby JV, Patient-Centered Outcomes Research I. The PCORI perspective on patient-centered outcomes research. *Jama*. 2014;312(15):1513-4.
147. Methodology Committee of the Patient-Centered Outcomes Research I. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *Jama*. 2012;307(15):1636-40.
148. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *Jama*. 2012;307(15):1583-4.
149. Gandhi GY, Murad MH, Fujiiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. *Jama*. 2008;299(21):2543-9.
150. Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. *BMJ*. 2010;341:c5707.
151. Gaudry S, Messika J, Ricard JD, Guillo S, Pasquet B, Dubief E, et al. Patient-important outcomes in randomized controlled trials in critically ill patients: a systematic review. *Ann Intensive Care*. 2017;7(1):28.
152. Qaseem A, Snow V, Cross JT, Jr., Forciea MA, Hopkins R, Jr., Shekelle P, et al. Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Ann Intern Med*. 2008;148(5):370-8.
153. Hewlett S, Wit M, Richards P, Quest E, Hughes R, Heiberg T, et al. Patients and professionals as research partners: challenges, practicalities, and benefits. *Arthritis Rheum*. 2006;55(4):676-80.
154. Ocana A, Tannock IF. When Are "Positive" Clinical Trials in Oncology Truly Positive? *JNCI: Journal of the National Cancer Institute*. 2010;103(1):16-20.
155. Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern Med*. 2013;173(8):611-2.
156. Naci H, Ioannidis JP. How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? Annual review of pharmacology and toxicology. 2015;55:169-89.
157. Gluud C, Krogsgaard K. Would you trust a surrogate respondent? *Lancet*. 1997;349(9053):665-6.
158. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. *N Engl J Med*. 1989;321(6):406-12.
159. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ*. 2013;346:f457.
160. Moynihan R. Surrogates under scrutiny: fallible correlations, fatal consequences. *BMJ*. 2011;343:d5160.

161. Walter SD, Sun X, Heels-Ansdell D, Guyatt G. Treatment effects on patient-important outcomes can be small, even with large effects on surrogate markers. *J Clin Epidemiol.* 2012;65(9):940-5.
162. Sautenet B, Caille A, Halimi JM, Goupille P, Giraudeau B. Better reporting and greater homogeneity in outcome measures are seen in randomized trial protocols when guidelines exist. *J Clin Epidemiol.* 2013;66(8):838-46.
163. Chong LSH, Sautenet B, Tong A, Hanson CS, Samuel S, Zappitelli M, et al. Range and Heterogeneity of Outcomes in Randomized Trials of Pediatric Chronic Kidney Disease. *J Pediatr.* 2017;186:110-7 e11.
164. Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med.* 2008;5(4):e96.
165. Johnston BC, Shamseer L, da Costa BR, Tsuyuki RT, Vohra S. Measurement issues in trials of pediatric acute diarrheal diseases: a systematic review. *Pediatrics.* 2010;126(1):e222-31.
166. Saldanha IJ, Li T, Yang C, Owczarzak J, Williamson PR, Dickersin K. Clinical trials and systematic reviews addressing similar interventions for the same condition do not consider similar outcomes to be important: a case study in HIV/AIDS. *J Clin Epidemiol.* 2017;84:85-94.
167. Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *Jama.* 2007;297(5):468-70.
168. Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ.* 2015;350:h809.
169. Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. *J Clin Epidemiol.* 2014.
170. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PMM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? *The Lancet.* 2016;387(10027):1573-86.
171. Chalmers I. Underreporting research is scientific misconduct. *Jama.* 1990;263(10):1405-8.
172. Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS ONE.* 2014;9(12):e114023.
173. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias G. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS ONE.* 2013;8(7):e66844.
174. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE.* 2008;3(8):e3081.
175. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess.* 2010;14(8):iii, ix-xi, 1-193.

176. Galsworthy MJ, Hristovski D, Lusa L, Ernst K, Irwin R, Charlesworth K, et al. Academic output of 9 years of EU investment into health research. *Lancet*. 2012;380(9846):971-2.
177. McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. Reporting bias in medical research - a narrative review. *Trials*. 2010;11:37.
178. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.gov: a cross-sectional analysis. *PLoS Med*. 2009;6(9):e1000144.
179. Martinez V, Pichard X, Fletcher D. Perioperative pregabalin administration does not prevent chronic postoperative pain: systematic review with a meta-analysis of randomized trials. *Pain*. 2017;158(5):775-83.
180. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess*. 2000;4(10):1-115.
181. Misakian AL, Bero LA. Publication bias and research on passive smoking: comparison of published and unpublished studies. *Jama*. 1998;280(3):250-3.
182. Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. *Jama*. 2003;290(4):495-501.
183. Jefferson T, Jones M, Doshi P, Del Mar C. Neuraminidase inhibitors for preventing and treating influenza in healthy adults: systematic review and meta-analysis. *BMJ*. 2009;339:b5106.
184. Godlee F. Open letter to Roche about oseltamivir trial data. *BMJ*. 2012;345:e7305.
185. Godlee F, Clarke M. Why don't we have all the evidence on oseltamivir? *BMJ*. 2009;339:b5351.
186. Abbasi K. The missing data that cost \$20bn. *BMJ : British Medical Journal*. 2014;348.
187. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov - The Final Rule. *N Engl J Med*. 2016.
188. Giannuzzi V, Altavilla A, Ruggieri L, Ceci A. Clinical Trial Application in Europe: What Will Change with the New Regulation? *Sci Eng Ethics*. 2016;22(2):451-66.
189. Nguyen TA, Dechartres A, Belgherbi S, Ravaud P. Public availability of results of trials assessing cancer drugs in the United States. *J Clin Oncol*. 2013;31(24):2998-3003.
190. Goldacre B, DeVito NJ, Heneghan C, Irving F, Bacon S, Fleminger J, et al. Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ*. 2018;362:k3218.
191. Page MJ, McKenzie JE, Forbes A. Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. *J Clin Epidemiol*. 2013;66(5):524-37.
192. Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR. Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database Syst Rev*. 2011(1):MR000031.

193. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *Jama*. 2009;302(9):977-84.
194. Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med*. 2005;24(10):1547-61.
195. Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC medicine*. 2015;13:282.
196. Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med*. 2008;5(11):e217; discussion e.
197. Chan AW, Krolez-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ*. 2004;171(7):735-40.
198. Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753.
199. Martinez V, Beloeil H, Marret E, Fletcher D, Ravaud P, Trinquart L. Non-opioid analgesics in adults after major surgery: systematic review with network meta-analysis of randomized trials. *Br J Anaesth*. 2017;118(1):22-31.
200. Bafeta A, Koh M, Riveros C, Ravaud P. Harms Reporting in Randomized Controlled Trials of Interventions Aimed at Modifying Microbiota: A Systematic Review. *Ann Intern Med*. 2018;169(4):240-7.
201. Derumeaux G, Ernande L, Serusclat A, Servan E, Bruckert E, Rousset H, et al. Echocardiographic evidence for valvular toxicity of benfluorex: a double-blind randomised trial in patients with type 2 diabetes mellitus. *PLoS ONE*. 2012;7(6):e38273.
202. Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? *BMJ*. 2007;334(7585):120-3.
203. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med*. 2000;343(21):1520-8, 2 p following 8.
204. Green LW, Ottoson JM, Garcia C, Hiatt RA. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annual review of public health*. 2009;30.
205. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *Journal of the Royal Society of Medicine*. 2011;104(12):510-20.
206. Trochim W, editor Translation won't happen without dissemination and implementation: Some measurement and evaluation issues. 3rd Annual Conference on the Science of Dissemination and Implementation Bethesda, MD: NIH Office of Behavioral and Social Sciences Research; 2010.
207. Westfall JM, Mold J, Fagnan L. Practice-based research—"Blue Highways" on the NIH roadmap. *Jama*. 2007;297(4):403-6.

208. Pathman DE, Konrad TR, Freed GL, Freeman VA, Koch GG. The awareness-to-adherence model of the steps to clinical guideline compliance. The case of pediatric vaccine recommendations. *Med Care*. 1996;34(9):873-89.
209. Glasziou P, Haynes B. The paths from research to improved health outcomes. *ACP J Club*. 2005;142(2):A8-10.
210. Simpson D, Sullivan GM. Knowledge Translation for Education Journals in the Digital Age. *J Grad Med Educ*. 2015;7(3):315-7.
211. Nickson CP, Cadogan MD. Free Open Access Medical education (FOAM) for the emergency physician. *Emergency medicine Australasia : EMA*. 2014;26(1):76-83.
212. Yordanov Y, Pateron D. Re: Not Another Boring Lecture: Let's Be the Guide on the Side. *Journal of Emergency Medicine*. 2015;49(5):710-1.
213. Thoma B, Chan T, Desouza N, Lin M. Implementing peer review at an emergency medicine blog: bridging the gap between educators and clinical experts. *CJEM*. 2015;17(2):188-91.
214. Pino C, Boutron I, Ravaud P. Inadequate description of educational interventions in ongoing randomized controlled trials. *Trials*. 2012;13:63.
215. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008;336(7659):1472-4.
216. Glenton C, Underland V, Kho M, Pennick V, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. *J Clin Epidemiol*. 2006;59(8):770-8.
217. Duff JM, Leather H, Walden EO, LaPlant KD, George TJ, Jr. Adequacy of published oncology randomized controlled trials to provide therapeutic details needed for clinical application. *J Natl Cancer Inst*. 2010;102(10):702-5.
218. Roberts I, Yates D, Sandercock P, Farrell B, Wasserberg J, Lomas G, et al. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. *Lancet*. 2004;364(9442):1321-8.
219. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, Farrell B, et al. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. *Lancet*. 2005;365(9475):1957-9.
220. collaborators C-t, Shakur H, Roberts I, Bautista R, Caballero J, Coats T, et al. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet*. 2010;376(9734):23-32.
221. collaborators C-, Roberts I, Shakur H, Afolabi A, Brohi K, Coats T, et al. The importance of early treatment with tranexamic acid in bleeding trauma patients: an exploratory analysis of the CRASH-2 randomised controlled trial. *Lancet*. 2011;377(9771):1096-101, 101 e1-2.
222. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *Jama*. 2010;303(20):2058-64.

223. Vera-Badillo FE, Napoleone M, Krzyzanowska MK, Alibhai SM, Chan AW, Ocana A, et al. Bias in reporting of randomised clinical trials in oncology. *Eur J Cancer*. 2016;61:29-35.
224. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol*. 2014;32(36):4120-6.
225. Liberati A. An unfinished trip through uncertainties. *BMJ*. 2004;328(7438):531.
226. UK health, science, and overseas aid: not what they seem. *Lancet*. 2010;376(9751):1437.
227. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *Jama*. 1999;282(11):1054-60.
228. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med*. 2006;3(10):e425.
229. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol*. 2004;57(6):543-50.
230. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *Jama*. 2002;287(21):2817-20.
231. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:MR000030.
232. Hoffmann TC, Erueti C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *BMJ*. 2013;347:f3755.
233. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.
234. Mhaskar R, Djulbegovic B, Magazín A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol*. 2012;65(6):602-9.
235. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*. 2004;328(7430):22-4.
236. Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol*. 2011;64(7):718-42.
237. Simera I. EQUATOR Network collates resources for good research. *BMJ*. 2008;337:a2471.
238. Larson EL, Cortazar M. Publication guidelines need widespread adoption. *J Clin Epidemiol*. 2012;65(3):239-46.
239. Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, Zhang S, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *Journal of multidisciplinary healthcare*. 2013;6:169-88.

240. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev*. 2012;1:60.
241. Koroshetz W. A core set of trial outcomes for every medical discipline? *BMJ*. 2015;350:h85.
242. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy*. 2012;17(1):1-2.
243. Tunis SR, Clarke M, Gorst SL, Gargon E, Blazeby JM, Altman DG, et al. Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*. 2016.
244. Murad MH, Shah ND, Van Houten HK, Ziegenfuss JY, Deming JR, Beebe TJ, et al. Individuals with diabetes preferred that future trials use patient-important outcomes and provide pragmatic inferences. *J Clin Epidemiol*. 2011;64(7):743-8.
245. Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. *BMC medicine*. 2016;14(1):100.
246. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. . In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated September 2011]. ed. Available from www.cochrane-handbook.org. : The Cochrane Collaboration; 2011.
247. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Akl E, et al. Chapter 11: Completing 'Summary of findings' tables and grading the confidence in or quality of the evidence. In: Higgins JPT, Chandler J, (editors) CM, editors. *Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0* (updated June 2017). ed. Cochrane, 2017. Available from <http://training.cochrane.org/handbook>: The Cochrane Collaboration; 2017.
248. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ. What is "quality of evidence" and why is it important to clinicians? *Bmj*. 2008;336(7651):995-8.
249. 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.
250. Ameer H, Ravaud P, Fayard F, Riveros C, Dechartres A. Systematic reviews of therapeutic interventions frequently consider patient-important outcomes. *J Clin Epidemiol*. 2017.
251. Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, Akl EA, Vernooij RW, et al. Systematic reviews experience major limitations in reporting absolute effects. *J Clin Epidemiol*. 2016;72:16-26.
252. Agarwal A, Johnston BC, Vernooij RW, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, et al. Authors seldom report the most patient-important outcomes and absolute effect measures in systematic review abstracts. *J Clin Epidemiol*. 2016.
253. Yordanov Y, Dechartres A, Ravaud P. Patient-important outcomes in systematic reviews: Poor quality of evidence. *PLoS ONE*. 2018;13(4):e0195460.
254. Yordanov Y, Dechartres A, Atal I, Tran VT, Boutron I, Crequit P, et al. Avoidable waste of research related to outcome planning and reporting in clinical trials. *BMC medicine*. 2018;16(1):87.

255. Ioannidis JP, Horbar JD, Ovelman CM, Brosseau Y, Thorlund K, Buus-Frank ME, et al. Completeness of main outcomes across randomized trials in entire discipline: survey of chronic lung disease outcomes in preterm infants. *BMJ*. 2015;350:h72.
256. Wuytack F, Smith V, Clarke M, Williamson P, Gargon E. Towards core outcome set (COS) development: a follow-up descriptive survey of outcomes in Cochrane reviews. *Syst Rev*. 2015;4(1):73.
257. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative [updated 2017 Available from: <http://www.comet-initiative.org/>.
258. Clarke M, Williamson P. Core outcome sets and trial registries. *Trials*. 2015;16(1):216.
259. 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.
260. Freedman B. Equipoise and the Ethics of Clinical Research. *New England Journal of Medicine*. 1987;317(3):141-5.
261. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781-8.
262. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390.
263. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *Jama*. 2013;309(8):814-22.
264. Montgomery P, Grant S, Hopewell S, Macdonald G, Moher D, Michie S, et al. Protocol for CONSORT-SPI: an extension for social and psychological interventions. *Implementation science : IS*. 2013;8:99.
265. Chan AW, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. *Lancet*. 2013;381(9861):91-2.
266. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Ann Intern Med*. 2013;158(3):200-7.
267. Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
268. Savović J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, et al. Development of a combined database for meta-epidemiological research. *Research Synthesis Methods*. 2010;1(3-4):212-25.
269. Nexus E. *Trials.ai Presentation*2016.

270. Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AH. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes*. 2011;4:304.
271. Tian L, Jiang F, Hasegawa T, Uno H, Pfeffer M, Wei LJ. Moving beyond the conventional stratified analysis to estimate an overall treatment efficacy with the data from a comparative randomized clinical study. *Statistics in Medicine*. 2018;0(0).
272. Galimard J-E, Chevret S, Protopopescu C, Resche-Rigon M. A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. *Statistics in Medicine*. 2016;35(17):2907-20.
273. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. (1471-2288 (Electronic)).
274. Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. (1873-622X (Electronic)).
275. Gabrio A, Mason AJ, Baio G. Handling Missing Data in Within-Trial Cost-Effectiveness Analysis: A Review with Future Recommendations. (2509-4254 (Electronic)).
276. Dembe AE, Partridge JS, Geist LC. Statistical software applications used in health services research: analysis of published studies in the U.S. *BMC Health Serv Res*. 2011;11:252.
277. BiostaTGV - Statistiques en ligne [Available from: <https://biostatgv.sentiweb.fr/>].
278. Barkan H. Statistics in clinical research: Important considerations. *Ann Card Anaesth*. 2015;18(1):74-82.
279. Na J, Yang H, Bae S, Lim KM. Analysis of Statistical Methods Currently used in Toxicology Journals. *Toxicol Res*. 2014;30(3):185-92.
280. Al-Benna S, Al-Ajam Y, Way B, Steintraesser L. Descriptive and inferential statistical methods used in burns research. *Burns*. 2010;36(3):343-6.
281. Wainapel SF, Kayne HL. Statistical methods in rehabilitation research. *Arch Phys Med Rehabil*. 1985;66(5):322-4.
282. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. *New England Journal of Medicine*. 2016;374(6):504-6.
283. Williamson P, Clarke M. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. *Cochrane Database Syst Rev*. 2012;5:ED000041.
284. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials*. 2012;13:132.
285. Gargon E, Williamson PR, Altman DG, Blazeby JM, Tunis S, Clarke M. The COMET Initiative database: progress and activities update (2015). *Trials*. 2017;18(1):54.
286. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET initiative database: progress and activities update (2014). *Trials*. 2015;16:515.

287. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities from 2011 to 2013. *Trials*. 2014;15:279.
288. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med*. 2017;14(11):e1002447.
289. Huwiler-Muntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *Jama*. 2002;287(21):2801-4.
290. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*. 2017;357:j2490.
291. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC medicine*. 2015;13:221.
292. Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ*. 2012;344:e4178.
293. Hopewell S, Boutron I, Altman DG, Barbour G, Moher D, Montori V, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. *BMC medicine*. 2016;14(1):199.
294. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract*. 2012;18(1):12-8.
295. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the Risk of Bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol*. 2012.
296. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339:b4012.
297. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ*. 2013;346:f1798.
298. Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing Risk of Bias in Randomised Clinical Trials Included in Cochrane Reviews: The why is Easy, the how is a Challenge [editorial]. *Cochrane Database Syst Rev*. 2013;5:ED000058.
299. Cochrane Methods Group. Risk of bias tools - RoB 2 tool [Available from: <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>].
300. Moseley AM, Elkins MR, Herbert RD, Maher CG, Sherrington C. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol*. 2009;62(10):1021-30.

301. Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. *Eval Health Prof.* 2002;25(1):116-29.
302. Goldkuhle M, Narayan VM, Weigl A, Dahm P, Skoetz N. A systematic assessment of Cochrane reviews and systematic reviews published in high-impact medical journals related to cancer. *BMJ Open.* 2018;8(3):e020869.
303. Gøtzsche PC, Ioannidis JPA. Content area experts as authors: helpful or harmful for systematic reviews and meta-analyses? *BMJ : British Medical Journal.* 2012;345.
304. Benger J. Getting our priorities right in emergency medicine. *Emergency medicine journal : EMJ.* 2017;34(7):457.
305. Smith JE, Morley R. The emergency medicine research priority setting partnership. *Emergency medicine journal : EMJ.* 2015;32(11):830.
306. Roberts I, Ker K. How systematic reviews cause research waste. *Lancet (London, England).* 2015;386(10003):1536.
307. Roberts I, Ker K. Cochrane: the unfinished symphony of research synthesis. *Syst Rev.* 2016;5(1):115.
308. Chevreton S, Ferguson ND, Bellomo R. Are systematic reviews and meta-analyses still useful research? No. *Intensive Care Med.* 2018;44(4):515-7.
309. Clark L, Fairhurst C, Hewitt CE, Birks Y, Brabyn S, Cockayne S, et al. A methodological review of recent meta-analyses has found significant heterogeneity in age between randomized groups. *Journal of Clinical Epidemiology.* 2014;67(9):1016-24.
310. Moore RA, Gavaghan D, Tramer MR, Collins SL, McQuay HJ. Size is everything--large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. *Pain.* 1998;78(3):209-16.

Quand on me dit qu'une personne est arrivée, je suis tentée de demander où, et par quels moyens, et dans quel but ! Seuls se félicitent ceux qui se savent incapables d'aller plus loin.

- Amin Maalouf -

