

**THESE DE DOCTORAT DE L'UNIVERSITE PARIS.DIDEROT (PARIS 7)
SORBONNE PARIS CITE**

Ecole doctorale Pierre Louis de sante publique à Paris

Epidémiologie et sciences de l'information biomédicale

Spécialité : Epidémiologie Clinique

Présenté par

AMELIE YAVCHITZ

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS DIDEROT PARIS VII

**Communication des Résultats Scientifiques dans les Essais Randomisés Contrôlés
et les Revues Systématiques**

Thèse dirigée par le Pr Isabelle BOUTRON

Présentée et soutenue publiquement le 13 Octobre 2015

Devant le jury composé de:

Monsieur Dominique Fletcher

Rapporteur

Monsieur Stefan Michiels

Rapporteur

Madame France Mentré

Examineur

Monsieur Bruno Falissard

Examineur

Madame Isabelle Boutron

Directeur de thèse

Remerciements

En tout premier lieu, je souhaite exprimer ma profonde gratitude à Monsieur le Professeur Philippe Ravaud qui m'a accueilli dans son équipe de recherche et m'a fait bénéficier durant mon Master et mon Doctorat de ses conseils avisés et de ses grandes qualités scientifiques.

Mes plus sincères remerciements vont à Madame le Professeur Isabelle Boutron, ma directrice de thèse. Je voudrais la remercier d'avoir cru en moi et en mes capacités. Je suis très heureuse qu'elle ait guidé mes premiers pas en recherche et qu'elle m'ait permis de travailler sur des sujets passionnants. J'espère que notre collaboration scientifique perdurera.

Je souhaite exprimer toute ma gratitude envers les membres du jury qui ont accepté de juger mon travail. J'adresse mes sincères remerciements à Monsieur le Professeur Dominique Fletcher et à Monsieur Stefan Michiels qui ont bien voulu être rapporteurs de ce travail, et à Madame le Professeur France Mentré et Monsieur le Professeur Bruno Falissard qui ont accepté de participer à ce jury de thèse.

Je souhaite remercier particulièrement Gabriel Baron et Isabelle Pane pour leur soutien statistique, informatique et webdesign qui ont permis la réalisation des travaux de cette thèse, ainsi qu'Aurélié, Pascal et Olivier pour leur relecture minutieuse.

Je tiens aussi à remercier tous les membres du Centre d'Epidémiologie clinique, de l'unité INSERM U1153, pour les bons moments passés ensemble en particulier Isabelle, Aïda, Agnès, Alexandre, Anne, Audrey, Aurélié, Aymar, Céline, Elise, Ludovic, Magalie, Mickael, Perrine, Raphaël, Romana, Thi et Youri.

Un très grand merci à ma famille, toujours présente, toujours unie, tant dans les moments joyeux que dans les moments difficiles.

A Olivier, Raphaël et Clémentine, les lumières de ma vie.

Résumé français

Les essais randomisés contrôlés et les revues systématiques avec méta-analyses sont les études considérées comme apportant le plus haut niveau de preuve ; elles sont devenues des éléments indispensables pour l'évaluation thérapeutique. Pour pouvoir interpréter au mieux les études et limiter les biais d'interprétation, les auteurs doivent rapporter de manière transparente et juste tous les éléments permettant aux lecteurs de faire leur propre analyse critique. Les *spin* sont des stratégies conscientes ou inconscientes utilisées par les auteurs pour « enjoliver » les résultats des études et sont potentiellement à l'origine d'un biais d'interprétation. D'une part, aucune étude ne s'est intéressée à la diffusion des *spin* et à leur impact sur l'interprétation des résultats par les lecteurs. D'autre part, plusieurs travaux se sont intéressés aux différents facteurs pouvant modifier l'interprétation des résultats de la recherche, mais aucun n'a évalué l'impact de la section « limites » dans les résumés de revues systématiques. Enfin, alors que les *spin* sont fréquents dans les publications scientifiques, aucune classification des *spin* dans les revues systématique et les méta-analyses n'a été proposée à ce jour.

Nos travaux ont montré 1) que les *spin* sont largement diffusés dans les communiqués et les articles de presse et qu'ils sont associés à une surestimation de l'effet bénéfique du traitement par les lecteurs, et 2) que l'ajout d'une section limite dans les résumés des revues systématiques et des méta-analyses ne modifie pas la confiance accordée par les lecteurs aux résultats de la revue. Enfin nous avons développé une classification des *spin* dans les revues systématiques et les méta-analyses et nous avons identifié les *spin* ayant la plus grande probabilité de fausser l'interprétation des lecteurs.

Mot clés : revues systématiques ; méta-analyses ; *spin* ; biais d'interprétation ; surestimation ; limites ; interprétation ; diffusion ; biais de diffusion ; essai randomisé contrôlé ; communication scientifique

Résumé en anglais

Randomized controlled trials (RCT) and systematic reviews and meta-analyses are the cornerstones of therapeutic evaluation because they are considered to provide the highest level of evidence. An accurate and complete presentation by authors of all the important information in report of such studies is essential to allow a critical appraisal of the study by readers. We define spin as a specific way of reporting, deliberate or not, that can lead to a “beautification” of the data. No study evaluated the dissemination of spin and their impact on the study’s interpretation. Furthermore, several studies focused on different elements that can affect the interpretation of research results, but none assessed the impact of adding a limitation section in abstract of systematic reviews. Finally, spin are frequent and classification was developed for spin in RCTs, however no classification of spin in systematic reviews and meta-analyses have been proposed.

Our work showed that 1) spin are disseminated in press releases and news and they are associated with an overestimation of the beneficial effect of treatment, 2) the addition of a limitation section in abstract of systematic reviews and meta-analysis does not impact the confidence in the study results by readers. Finally, we developed a classification scheme of spin in systematic reviews and meta-analyses, and we ranked spin in abstract according to their perceived severity (i.e. the likelihood to distort the interpretation of the review).

Keywords : systematic review ; meta-anlaysis ; interpretation bias ; overestimation ; limits ; interpretation ; dissimulation ; dissemination bias

Laboratoire d'accueil

Centre de Recherche Epidémiologie et Statistiques Sorbonne Paris Cité (CRESS)

CRESS U1153, équipe METHODS : Méthodes en Évaluation Thérapeutique des Maladies Chroniques

Centre d'Épidémiologie Clinique

Hôpital Hôtel Dieu

1 place du parvis Notre Dame

75181 Paris, France

Financement :

Obtention de deux bourses de thèse

2011 : Fonds d'Etudes et de Recherche du Corps Médical des Hôpitaux de Paris

2014 : Fondation pour la Recherche Médicale, FDT20140930875

Publications liées à la thèse

Ce travail de doctorat s'appuie sur les articles suivants :

Articles publiés

Yavchitz A, Ravaud P, Hopewell S, Baron G, Boutron I.

Impact of adding a limitations section to abstracts of systematic reviews on readers' interpretation: a randomized controlled trial.

BMC Med Res Methodol. 2014 Nov 24;14(1):123. doi: 10.1186/1471-2288-14-123.

Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, Ravaud P.

Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study.

PLoS Med. 2012;9(9):e1001308. doi: 10.1371/journal.pmed.1001308. Epub 2012 Sep 11.

Articles en cours

Yavchitz A, Ravaud P, Moher D, Altman D, Lasserson T, Hrobjartsson A, Boutron I.

Classification and Ranking of Spin in Systematic Reviews and Meta-analyses.

Table des Matières

1. INTRODUCTION ET OBJECTIFS	14
1.1. DE LA RECHERCHE A LA PRATIQUE CLINIQUE	15
1.1.1. <i>La médecine fondée sur les preuves</i>	15
1.1.1.1. La pratique de l'EBM	15
1.1.1.2. La force de la preuve	16
1.1.1.3. Les revues systématiques et les méta-analyses.....	17
Revue systématique et méta-analyse : des outils devenus indispensables	17
Historique.....	20
Les grands principes des revues systématiques et des méta-analyses	21
1.1.2. <i>Diffusion des résultats de la recherche</i>	24
1.1.2.1. Les publications scientifiques : rôle majeur du résumé d'article	24
1.1.2.2. Les congrès scientifiques.....	25
1.1.2.3. La diffusion médiatique des résultats de la recherche	26
1.2. INTERPRETATION DES RESULTATS DE LA RECHERCHE.....	28
1.2.1. <i>Les difficultés d'interprétation des résultats</i>	28
1.2.2. <i>Les facteurs pouvant influencer l'interprétation des résultats par les lecteurs</i>	30
1.2.2.1. Les croyances antérieures	30
1.2.2.2. Les conflits d'intérêt	32
1.2.2.3. Les modes de représentation des résultats	33
1.2.3. <i>Transparence</i>	36
1.2.3.1. Les reporting guidelines	36
1.2.3.2. Les registres d'essais.....	38
1.3. BIAIS DE DIFFUSION DES RESULTATS.....	40
1.3.1. <i>Le biais de publication des articles</i>	41
1.3.2. <i>La présentation sélective des critères de jugement</i>	43
1.3.3. <i>Les spin</i>	45

1.3.3.1.	Définition du concept de <i>spin</i>	45
1.3.3.2.	Les <i>spin</i> dans les essais randomisés contrôlés	46
1.3.3.3.	Les <i>spin</i> dans les études diagnostiques	51
1.3.3.4.	Les <i>spin</i> dans les études observationnelles et les revues systématiques.....	52
1.3.3.5.	Prévalence des <i>spin</i>	54
1.4.	JUSTIFICATION DE LA DEMARCHE DE THESE	56
2.	INTERPRETATION DES RESULTATS DES ESSAIS RANDOMISES CONTROLES DANS LES COMMUNIQUES DE PRESSE ET LES ARTICLES DE PRESSE : UNE ETUDE DE COHORTE.	58
2.1.	JUSTIFICATION DE L'ETUDE.....	59
2.2.	OBJECTIF	59
2.3.	METHODE.....	59
2.4.	RESULTATS	62
2.5.	DISCUSSION.....	64
2.6.	ARTICLE.....	66
3.	IMPACT DE L'AJOUT D'UNE SECTION LIMITE DANS LES RESUMES DES REVUES SYSTEMATIQUES SUR L'INTERPRETATION DES LECTEURS.....	77
3.1.	JUSTIFICATION DE L'ETUDE.....	78
3.2.	OBJECTIF	78
3.3.	METHODE.....	78
3.4.	RESULTATS	80
3.5.	DISCUSSION.....	83
3.6.	ARTICLE.....	85
4.	CLASSIFICATION DES SPIN DANS LES REVUES SYSTEMATIQUES ET LES META-ANALYSES.....	95
4.1.	JUSTIFICATION DE L'ETUDE.....	96
4.2.	OBJECTIFS	96

4.3.	METHODES.....	97
4.3.1.	<i>Définitions.....</i>	97
4.3.2.	<i>Développement de la classification des spin dans les revues systématiques et les méta-analyses.....</i>	97
4.3.3.	<i>Classement des spin dans les résumés des revues systématiques et méta-analyses.....</i>	98
4.4.	RESULTATS	100
4.4.1.	<i>Identification des spin dans les revues systématiques et les méta-analyses.....</i>	100
4.4.2.	<i>Classification des spin dans les revues systématiques et les méta-analyses</i>	101
4.4.3.	<i>Résultats de l'enquête sur la sévérité des spin dans les résumés des revues systématiques et méta-analyses.....</i>	102
4.5.	DISCUSSION.....	107
4.6.	ARTICLE.....	109
5.	DISCUSSION ET PERSPECTIVES	138
5.1.	DISCUSSION.....	139
5.1.1.	<i>Principaux résultats.....</i>	139
5.1.2.	<i>Place de nos travaux dans la recherche.....</i>	140
5.1.3.	<i>Implications.....</i>	142
5.1.3.1.	<i>Le rôle de la communauté scientifique dans la diffusion des spin.....</i>	142
5.1.3.2.	<i>Améliorer la communication des résultats scientifiques: un besoin de nouveaux outils</i>	143
5.2.	CONCLUSION	147
5.3.	PERSPECTIVES	149
5.3.1.	<i>Fréquence, diffusion et impact des spin dans les revues systématiques et les méta-analyses.....</i>	149
5.3.2.	<i>Développement d'outils d'aide à la prévention et à la détection des spin.....</i>	150
5.3.3.	<i>De nouveaux outils pour améliorer la communication des résultats des revues systématiques.....</i>	151

BIBLIOGRAPHIE.....	153
ANNEXES.....	168
<i>Annexe 1 : Stratégies de recherche</i>	<i>168</i>
<i>Annexe 2 : Mails d'invitation et copies d'écran du site de l'étude.....</i>	<i>169</i>
<i>Annexe 3 : Mail d'invitation au Q-Sort</i>	<i>173</i>
<i>Annexe 4: Répartition des réponses pour chaque item.....</i>	<i>174</i>

Listes des Figures

Figure 1: Les différents niveaux de preuve.....	17
Figure 2 : Différentes représentations du même effet traitement.....	35
Figure 3: Les étapes de l'enquête en ligne.....	100
Figure 4: Les différences de description des éléments méthodologiques entre 2000 et 2006 [180]	144

Liste des tableaux

Tableau 1: Décisions des huit experts à partir des six méta-analyses qui leur sont présentées (Méta-analyses cumulatives) [51]	29
Tableau 2: Classification des <i>spin</i> dans les ERC	50
Tableau 3: Classification des <i>spin</i> dans les études de précision diagnostique	52
Tableau 4: Classification des <i>spin</i>	61
Tableau 5: Exemple de résumé avec et sans section limite.....	82
Tableau 6: Résultats du critère principal et des critères secondaires	83
Tableau 7 : Classification des <i>spin</i> dans les revues systématiques et les méta-analyses.	103
Tableau 8: Classement des <i>spin</i> selon leur sévérité	106

Abréviations

EBM: Evidence Based Medicine

PICO: Population - Intervention – Comparator - Outcome

ERC: Essai Randomisé Contrôlé

IC : Intervalle de Confiance

FDA : Food and Drug Administration

1. Introduction et Objectifs

1.1. De la recherche à la pratique clinique

1.1.1. La médecine fondée sur les preuves

La pratique de l'exercice médical a évolué au cours du 19^{ème} siècle. Fondé initialement sur une approche empirique, l'exercice médical s'est enrichi d'une approche plus scientifique, basé sur les résultats de l'expérimentation. Cette médecine factuelle, ou « evidence based medicine » (EBM), fait partie intégrante de la pratique médicale moderne. L'EBM consiste à prendre des décisions médicales fondées 1) sur les preuves scientifiques qui sont les connaissances issues de la recherche clinique avec une démarche systématique, rigoureuse et reproductible et 2) sur la connaissance théorique, l'expérience et le jugement clinique. La transposition des résultats de la recherche à la pratique clinique est donc un des fondements sur lesquels l'EBM repose.

1.1.1.1. La pratique de l'EBM

Le concept d'EBM fut initialement développé par des épidémiologistes canadiens de la McMaster Medical School (Ontario, Canada) dans les années 1980. Il s'agissait d'abord d'une méthode d'enseignement : la recherche par la preuve était utilisée comme méthode de raisonnement et source d'information pour la formation des étudiants en médecine. A partir des années 1990, l'EBM est devenue une méthodologie à l'usage des praticiens, afin de les aider dans leur prise de décision.

La pratique de l'EBM se déroule en 4 étapes :

1. Formuler une question claire et précise qui doit répondre aux critères PICO (acronyme de *Population – Intervention – Control – Outcome*).
2. Chercher dans la littérature les ressources pertinentes en rapport avec la question posée.
3. Réaliser l'évaluation critique de la validité et de la pertinence des résultats trouvés.
4. Intégrer les résultats obtenus dans la pratique clinique courante pour adapter au mieux la décision médicale à chaque patient.

La pratique de l'EBM implique d'une part d'avoir accès à l'ensemble de la littérature scientifique. Or le nombre de publications est considérable et ne cesse d'augmenter ; ce qui devient un obstacle majeur à l'intégration de l'EBM dans la pratique quotidienne. D'autre part, la pratique de l'EBM implique d'être capable d'utiliser et d'analyser de manière critique les ressources disponibles. Il est donc nécessaire de pouvoir hiérarchiser les publications en fonction de leur capacité à répondre à la question posée, ce que l'on définit comme étant le niveau de preuve d'une étude, ou force de preuve.

1.1.1.2. La force de la preuve

La force de la preuve, « level of evidence », est décrite pour la première fois en 1979 dans un rapport de la Canadian Task Force on the Periodic Health Examination (CTFPHE). Une méthodologie est alors développée afin de pouvoir hiérarchiser les différents niveaux de preuve apportés aux recommandations de pratique clinique. L'essai randomisé contrôlé (ERC) est considéré comme apportant le plus haut niveau de preuve car il s'agit du schéma d'étude qui contrôle le mieux les biais[1]. De ce fait, les recommandations qui s'appuient sur les résultats d'ERC ont une force de preuve plus importante que les recommandations basées sur des études observationnelles ou les avis d'experts. La première classification des niveaux de

preuve a été décrite par DL Sackett en 1989[2].

Cette classification est reprise et améliorée dans les années 1990 en intégrant les études qui s'appuient sur une synthèse méthodique de la littérature. Ces études appelées revues systématiques peuvent inclure ou non une synthèse quantitative des résultats appelé méta-analyse. Actuellement les revues systématiques avec méta-analyses d'ERC sont considérées comme étant les études apportant potentiellement le plus haut niveau de preuve. La figure 1 représente les différents niveaux de preuve selon T. Grenhalgh[3].

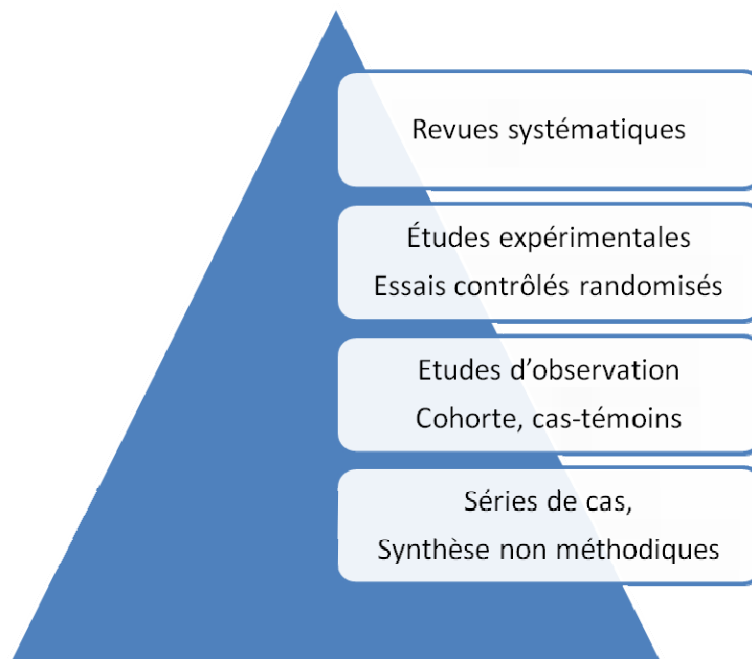


Figure 1: Les différents niveaux de preuve.

1.1.1.3. Les revues systématiques et les méta-analyses

Revue systématique et méta-analyse : des outils devenus indispensables

Initialement l'ERC était le schéma d'étude considéré comme ayant le plus haut niveau de

preuve. Chaque essai était analysé individuellement par le lecteur. Celui-ci devait faire l'effort de se tenir au fait des nouvelles publications et ainsi faire une synthèse critique de la littérature. Le lecteur se confronte alors à plusieurs difficultés : 1) une contrainte de temps, 2) l'analyse méthodologique de chaque étude et 3) l'accès aux publications.

En effet, il est devenu impossible pour les professionnels de santé de se tenir informé de tous les nouveaux articles publiés, de les lire et d'en faire une synthèse critique. D'une part cela nécessiterait un temps considérable et d'autre part cela supposerait d'avoir accès à l'ensemble de cette information et d'être capable de l'analyser de manière critique. Ainsi, en 2010, H. Bastian a évalué le nombre d'essais publiés annuellement, celui-ci passe d'environ 10 000 en 1990 à plus de 25 000 en 2007. En 2010, 75 essais étaient publiés quotidiennement et, entre le 1^{er} janvier et le 31 décembre 2014, plus de 15 000 ERC sont référencés sur MEDLINE[4].

Aussi, les essais ne doivent-ils pas être analysés de manière isolée. En effet, en 1904, K. Pearson soulignait déjà les limites de l'analyse des études de manière séparée. « 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 »[5]. L'absence de synthèse peut être à l'origine d'un retard dans l'élaboration des recommandations. Il s'écoule parfois plusieurs années entre la publication d'études montrant l'efficacité de nouvelles interventions et leur utilisation en pratique clinique. De même des interventions inefficaces voire nocives pour les patients continuent d'être recommandées malgré des études en leur défaveur[6]. Le besoin de synthèse est donc primordial et les revues systématiques jouent ce rôle[7].

Les revues systématiques ont plusieurs rôles[8]. Premièrement, elles permettent la généralisation des résultats scientifiques : les différentes études tentent de répondre à une même question mais dans des contextes parfois différents, avec des populations (plus ou

moins) différentes, des méthodes de mesure différentes, etc. Deuxièmement, elles permettent d'évaluer la cohérence des résultats et leur homogénéité ; en effet, la revue systématique permet de vérifier que les effets retrouvés au sein des différentes études sont toujours dans le même sens et ont le même ordre de grandeur. Troisièmement, elles permettent d'explorer les causes d'hétérogénéité dans les résultats. L'hétérogénéité peut être notamment d'origine clinique ; certaines stratégies thérapeutiques efficaces pour un type de patients peut s'avérer inefficaces voire délétères dans une autre population de malades. Aussi les résultats entre les études peuvent varier pour des raisons indépendantes de la population d'étude mais pour des raisons méthodologiques. Et enfin, les revues systématiques quantitatives, c'est-à-dire accompagnées d'une méta-analyse, permettent d'augmenter la puissance et la précision de l'effet estimé.

Les revues systématiques et les méta-analyses sont ainsi devenues des outils indispensables. Elles servent de support pour la formation continue des médecins et pour l'élaboration des recommandations de pratique clinique par les agences de santé et les sociétés savantes, ainsi que pour la prise de décision pour les politiques de santé.

Il a également été souligné qu'éthiquement et scientifiquement, tout nouvel essai clinique devrait être réalisé à la lumière de l'ensemble des données existantes, et donc idéalement après avoir réalisé une revue systématique[7,9]. Selon V. Roloff, un nouvel essai ne devrait être entrepris que si le résultat de la méta-analyse n'est pas concluant et si ce nouvel essai est d'une puissance suffisante pour changer les résultats de la méta-analyse mise à jour[10]. Dans la série du Lancet sur le gâchis évitable de la recherche (« avoidable waste of research ») publié en 2014[11–13], les auteurs ont souligné l'importance des revues systématiques pour éviter les recherches inutiles et définir les besoins. Avoir recours aux

revues systématiques éviterait d'inclure inutilement des patients dans des essais qui ne modifieront pas les données déjà acquises, d'autant plus que ces patients risqueraient alors de recevoir un traitement non optimal au vu des données de la science. Ainsi, toute nouvelle étude devrait commencer et finir par une revue systématique avec méta-analyse[14].

Historique

Il faudra attendre la fin des années 1970 pour que les revues systématiques et les méta-analyses se développent en médecine. Le médecin anglais Archibald Lemman Cochrane (1909-1988) fut l'un des pionniers de l'épidémiologie moderne. Il publia dans les années 1970 le premier article à être qualifié de revue systématique. Dans son ouvrage intitulé « Effectiveness and Efficiency » publié en 1972, AL Cochrane exposa sa théorie “that health services should be evaluated on the basis of scientific evidence rather than on clinical impression, anecdotal experience, 'expert' opinion or tradition”. Ce texte est considéré comme l'un des ouvrages fondateurs de la médecine factuelle[15]. En 1979, celui-ci attira l'attention de la communauté scientifique sur la nécessité de disposer de résumés régulièrement mis à jour et rapportant l'efficacité des interventions médicales : “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”[16].

La création de la Collaboration Cochrane en 1993 par Sir Ian Chalmers a fortement contribué au développement des revues systématiques et méta-analyses. Il s'agit d'une organisation à but non lucratif indépendante de l'industrie pharmaceutique. Elle produit et diffuse des revues systématiques sur l'évaluation des interventions en santé dans tous les domaines médicaux et

paramédicaux selon une méthodologie rigoureuse et transparente ainsi que la mise à jour et la diffusion de ces revues. Les membres de la Collaboration Cochrane sont impliqués dans des activités connexes comme: 1) favoriser les données probantes dans la prise de décisions en santé, 2) former à la préparation de revues Cochrane, 3) améliorer la méthodologie des revues systématiques et 4) traduire les revues Cochrane de l'anglais vers d'autres langues. La Collaboration Cochrane met à disposition un manuel de référence (le Cochrane Handbook [17]) et un logiciel (Revman).

Les grands principes des revues systématiques et des méta-analyses

Contrairement à une revue narrative, la revue systématique est une recherche rigoureuse permettant de produire une synthèse de l'ensemble des rapports médicaux existant dans un domaine médical précis. Les revues systématiques suivent un protocole explicite, précis et reproductible, associé à une analyse critique. Les méthodes rigoureuses explicites utilisées dans les revues systématiques permettent en théorie de limiter les biais, et donc d'augmenter la fiabilité et la précision des résultats[18,19].

La réalisation d'une revue systématique comporte 5 étapes essentielles qui sont :

- 1) La définition des objectifs, c'est-à-dire une question clinique claire et précise.
- 2) La définition des critères d'éligibilité c'est-à-dire la description des caractéristiques des études à inclure (population, critère de jugement, durée du suivi) et les critères utilisés pour sélectionner les études (année, langue et statut de la publication).
- 3) Une recherche systématique qui vise à identifier toutes les études (publiées ou non) qui

répondent aux critères d'éligibilité.

4) Une évaluation de la qualité des études incluses, c'est-à-dire de leur risque de biais.

5) Une présentation systématique avec la synthèse des caractéristiques et des résultats des études incluses.

Les méta-analyses synthétisent l'ensemble des résultats des études répondant à une question thérapeutique donnée, dans le but d'obtenir une estimation moyenne (combinée) de l'effet traitement.

Il existe deux types de méta-analyse : les méta-analyses réalisées sur les données résumées extraites des articles rapportant les résultats des études incluses et les méta-analyses sur données individuelles réalisées à partir des données individuelles des patients. Les méta-analyses calculent une moyenne pondérée de l'estimation de l'effet traitement, issue de chaque étude sélectionnée. Dans la méta-analyse, les études ayant une variance de l'estimation de l'effet traitement la plus petite ont un poids plus important et contribuent donc davantage à l'estimation de l'effet traitement moyen. Les deux principaux modèles statistiques qui existent pour combiner les résultats de plusieurs études sont le modèle à effet fixe et le modèle à effet aléatoire[20]. Dans un modèle à effet fixe, on considère que l'effet traitement est une constante. Chaque étude mesure cette même constante, ainsi la série d'études est une série de mesures de ce même effet traitement soumis à des fluctuations d'échantillonnage. Dans un modèle à effet aléatoire, on émet l'hypothèse que le résultat d'une étude est composé d'une partie fixe, commune à l'ensemble des études, et d'une partie propre à cette étude donc variable.

L'estimation de l'effet traitement provenant d'une méta-analyse réalisée dans le cadre d'une

revue systématique est moins biaisée que l'estimation de l'effet traitement provenant d'une étude isolée ou d'une méta-analyse non réalisée dans le cadre d'une revue systématique[21].

1.1.2. Diffusion des résultats de la recherche

Entre la fin d'une étude et la mise en pratique de ces résultats, le processus est long et complexe. Ce processus comporte de nombreuses étapes dont la première consiste à porter à la connaissance du plus grand nombre les résultats de l'étude. Les publications scientifiques dans des revues médicales, les congrès scientifiques et les médias sont les principaux vecteurs de l'information.

1.1.2.1. Les publications scientifiques : rôle majeur du résumé d'article

L'article scientifique a pour objectif de rapporter les résultats d'une étude, expérimentale ou non. Sa rédaction suit des règles rigoureuses dictées par les éditeurs. Le comité international des éditeurs de revues médicales (ICMJE) publie des recommandations, régulièrement mises à jour, pour l'écriture des articles scientifiques. L'ICMJE recommande que le texte d'un article suive toujours le même plan : IMRAD pour *Introduction, Methods, Results And Discussion*. Les articles originaux requièrent en plus du texte de l'article un résumé de celui-ci.

Le résumé d'article est probablement la section la plus importante d'un article scientifique pour plusieurs raisons. Premièrement, le résumé est la seule partie de la publication accessible à tous. Malgré l'essor des revues médicales en libre accès près de la moitié des publications scientifiques sont payantes ; en 2010, 50% des articles médicaux étaient accessibles gratuitement[22]. Par contre, dans les bases de données, les résumés sont toujours accessibles gratuitement et en anglais. En effet, les articles originaux sont indexés avec le titre et le résumé de l'article dans les bases de données comme MEDLINE ou EMBASE.

Deuxièmement, la première évaluation d'un article est faite à la lecture de son résumé et donnera l'envie ou non au lecteur de lire le texte intégral. Aussi, certains médecins ne lisent que les résumés d'article et se fient aux résultats rapportés dans le résumé [23,24]. Cela signifie que, parfois, des décisions médicales peuvent être prises sur la base de résultats rapportés dans le résumé plutôt que dans la publication intégrale.

C'est pourquoi le résumé mérite une attention particulière et fait l'objet de recommandations spécifiques[25,26]. De nombreuses politiques éditoriales ont été mises en œuvre pour améliorer le contenu et le format des résumés d'articles. Dans les années 1960, les résumés, habituellement placés à la fin de l'article, ont été déplacés au début[27], et les années 1980 ont vu l'essor des résumés structurés [28]. Malgré des améliorations [29], la qualité du *reporting* des résumés d'articles reste médiocre [30][31][32][33][34]. Par exemple, dans un échantillon de 182 résumés de revues systématiques, la direction de l'effet n'est pas clairement décrite dans 77 cas (42%), et dans 43 cas (24%) il n'est pas possible de déterminer de façon fiable la direction de l'effet estimé [31].

1.1.2.2. Les congrès scientifiques

Les congrès scientifiques sont une source d'information importante pour les professionnels de la santé. De nombreuses études sont présentées initialement sous forme de résumé lors des congrès ou conférences, et seront ensuite publiées dans les actes du congrès, parfois accessibles en ligne. En revanche ces résumés présentés en congrès ne sont pas indexés dans les bases de données type MEDLINE. Ces résumés peuvent fournir des informations précieuses pour la réalisation des revues systématiques car de nombreuses études ne seront

pas publiées, et les exclure pourrait introduire un biais[35].

Cependant, plusieurs études ont montré qu'il existait des différences importantes entre les données rapportées dans les résumés des actes de congrès par rapport à celles rapportées dans la publication ultérieure quand elle a lieu. Ces études questionnent la fiabilité et la qualité des informations rapportées dans les résumés présentés en congrès[36][30].

1.1.2.3. La diffusion médiatique des résultats de la recherche

Les médias jouent un rôle important dans la diffusion des résultats de la recherche[37] ; en effet 90% de la population s'informent sur les avancées scientifiques et médicales via les médias généralistes[38]. Les informations sur la santé publiées par les médias intéressent un large public et peuvent modifier le comportement des prescripteurs et des consommateurs de soins[39,40]. Par exemple, la diffusion des reportages concernant le dépistage par test rapide du streptocoque du groupe A(SGA) dans les services d'urgence pédiatrique a été associées à une augmentation importante de la prescription de ces tests, et ce malgré l'absence d'augmentation du nombre d'enfants présentant des symptômes qui pourraient les justifier [41]. Aussi, selon l'Institut de Recherche en Santé du Canada, près de 9 Canadiens sur 10 prennent directement des décisions concernant leur santé suite aux informations diffusées par les médias[42].

Les communiqués de presse sont au centre du processus de diffusion des résultats de la recherche. Les communiqués de presse scientifiques sont des textes courts destinés aux

journalistes dans le but de les informer d'un nouvel essai. Ils ont pour but d'être repris ou de donner envie aux journalistes de développer l'information dans un article. Il existe des bases de données spécialisées dans la diffusion des communiqués de presse scientifiques dont la plus importante est EurekaAlert!. Les communiqués de presse sont largement diffusés par les chercheurs, les institutions ou les compagnies pharmaceutiques pour attirer l'attention des médias et promouvoir leurs travaux de recherche[43–45].

Plusieurs auteurs ont étudié la qualité des communiqués et des articles de presse. Ils ont montré que la description des résultats des études était souvent inappropriée et qu'elle ne contenait pas les éléments indispensables à une bonne interprétation par les lecteurs[46–48].

Dans une cohorte de 200 communiqués de presse issus de dix facultés de médecine américaines, S. Woloshin a montré que les communiqués de presse favorisaient des études peu pertinentes pour l'homme (44% rapportaient des études animales ou de recherche fondamentale) et que les principales limites des études n'étaient pas rapportées[49]. Les communiqués et articles de presse ont tendance à amplifier l'importance des découvertes scientifiques. Par exemple quand un communiqué de presse extrapole les résultats issus de la recherche animale à l'homme, ou ne précise pas le nombre de patients inclus dans l'étude surtout quand il est très faible, peut faire croire à tort que le traitement étudié est efficace [48]. Par ailleurs dans une étude de L. Schwartz publiée en 2012 dans le *BMJ*, les communiqués de presse de bonne qualité émis par des revues médicales semblent être à l'origine d'articles de journaux de meilleure qualité, alors que les communiqués de presse de faible qualité sont associés à des articles de presse de mauvaise qualité[50].

1.2. Interprétation des résultats de la recherche.

1.2.1. Les difficultés d'interprétation des résultats.

Un des enjeux majeurs pour assurer la prise en charge des patients est le transfert des résultats de la recherche vers la pratique clinique courante. Pour cela il est important de comprendre le processus d'interprétation des résultats de la recherche et les facteurs pouvant modifier l'interprétation.

Alors que le dicton populaire dit que « les chiffres parlent d'eux-mêmes », de nombreux travaux nous montrent que l'interprétation des données de l'expérimentation scientifique est un phénomène complexe et en partie subjectif.

En effet, une même étude peut donner lieu à une interprétation différente et à des conclusions différentes de la part des auteurs comme des lecteurs. Ainsi, dans l'étude de I Shrier, huit relecteurs expérimentés en cardiologie ou en épidémiologie répondent à une série de questions sur l'efficacité pour les critères de mortalité et d'arythmie d'un traitement par magnésium intraveineux après un infarctus du myocarde. Les huit experts ont accès aux mêmes données issues d'une méta-analyse incluant 23 ERC et 69505 patients. A la question « I recommend that magnesium therapy be used in patients during the post-MI period » les huit experts ont des réponses différentes, puisque quatre experts répondent « oui » et quatre répondent « non »[51]. Voir tableau 1.

Tableau 1: Décisions des huit experts à partir des six méta-analyses qui leur sont présentées (Méta-analyses cumulatives) [51]

# RCTS	1	1-3	1-5	1-10	1-20	1-23
N	111	415	597	3685	63047	69505
Fixed OR	N/a	0.40 (0.19-0.83)	0.40 (0.28-0.61)	0.64 (0.52-0.79)	1.02 (0.96-1.08)	1.01 (0.96-1.07)
Rand OR	N/a	0.40 (0.18-0.86)	0.38 (0.21-0.66)	0.66 (0.53-0.81)	0.65 (0.48-0.87)	0.75 (0.61-0.92)
I ²	N/a	0%	0%	21%	59%	59%
I believe magnesium has now been shown to be beneficial for patients during the post-MI period						
(C)	Disagree	Unsure	Unsure	Disagree	Disagree	Disagree
(C)	Strongly Disagree	Unsure	Unsure	Unsure	Disagree	Strongly Disagree
(P)	Strongly Disagree	Agree	Agree	Agree	Agree	Agree
(P)	Agree	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Agree
(P)	Strongly Disagree	Strongly Disagree	Disagree	Unsure	Agree	Agree
(P)	Unsure	Agree	Agree	Agree	Agree	Agree
(NP)	Strongly Disagree	Disagree	Disagree	Agree	Disagree	Strongly Disagree
(N)	Unsure	Agree	Agree	Agree	Agree	Agree
I believe magnesium will eventually be shown to be beneficial for patients during the post-MI period						
(C)	Agree	Agree	Agree	Agree	Unsure	Disagree
(C)	Unsure	Unsure	Unsure	Unsure	Strongly Disagree	Strongly Disagree
(P)	Unsure	Agree	Agree	Strongly Agree	Agree	Agree
(P)	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Agree
(P)	Strongly Disagree	Strongly Disagree	Disagree	Unsure	Agree	Agree
(P)	Agree	Agree	Agree	Agree	Agree	Agree
(NP)	Unsure	Unsure	Unsure	Agree	Unsure	Disagree
(N)	Agree	Agree	Agree	Agree	Agree	Agree
I recommend that magnesium therapy be used in patients during the post-MI period						
(C)	No	No	No	No	No	No
(C)	No	No	No	No	No	No
(P)	No	Yes	Yes	Yes	Yes	Yes
(P)	No	Yes	Yes	Yes	Yes	Yes
(P)	No	No	No	No	No	No
(P)	No	Yes	Yes	Yes	Yes	Yes
(NP)	No	No	No	Yes	No	No
(N)	No	Yes	Yes	Yes	Yes	Yes

C: cardiologist, P: other physician, NP: non-practicing physician, N: non-physician

De même, dans l'étude de N.M. Lai[52], il était demandé à des étudiants en 5^{ème} année ainsi qu'à des médecins hospitaliers ayant suivi une formation à l'EBM, de choisir la conclusion la plus appropriée aux résultats de quatre revues systématiques avec méta-analyses de la Cochrane et dont la conclusion des auteurs avait été masquée. Dans cette étude, 60% des participants identifiaient correctement la direction de l'effet et seul 30% avaient choisi la conclusion la plus adaptée aux résultats de la revue en termes de direction de l'effet et de force de la preuve.

Les différences d'interprétation sont attribuées à une analyse et à un jugement différents quant à la stratégie de recherche, aux choix des critères d'exclusion ou d'inclusion des études

princeps ou aux choix des critères de jugement rapportés dans l'étude[53]. Les résultats de ces deux études montrent que, malgré une formation spécifique à l'EBM et un certain niveau d'expertise médicale ou méthodologique, il est difficile d'interpréter correctement une revue systématique et de générer une conclusion appropriée. Ces phénomènes sont responsables d'une interprétation potentiellement biaisée des résultats à l'origine d'un message erroné retenu par les lecteurs.

1.2.2. Les facteurs pouvant influencer l'interprétation des résultats par les lecteurs

L'étude de la perception de résultats est l'évaluation de la façon dont les individus perçoivent et comprennent des informations et dont ces informations sont utilisées dans les processus décisionnels. La perception de l'information a été largement décrite en psychologie pour comprendre les mécanismes cognitifs de perception du risque face à un danger réel ou supposé[54]. En médecine, la perception de l'information a été étudiée pour comprendre le processus de prise de décision notamment en chirurgie ou en anesthésie[55,56]. Dans notre domaine d'intérêt, les facteurs pouvant influencer l'interprétation sont soit liés aux lecteurs eux-mêmes (croyances antérieures, perception des conflits d'intérêt), soit liés au mode de présentation.

1.2.2.1. Les croyances antérieures

Les individus ont naturellement tendance à privilégier les informations qui confirment leurs

idées et à accorder moins de valeur aux données qui vont à l'encontre de leurs convictions ou croyances antérieures. En recherche, l'influence des croyances antérieures sur la perception des résultats est appelée le biais de confirmation, il désigne la tendance des lecteurs à accorder plus de crédit aux études dont les résultats confirment une conviction ou une connaissance antérieure et a contrario à juger les études dont les résultats sont en désaccord avec leurs convictions comme étant de moins bonne qualité.

Différents auteurs ont mis en évidence l'influence des croyances antérieures sur l'interprétation des résultats grâce à des études expérimentales [52,57,58].

J. Koehler a demandé à 297 étudiants en médecine d'évaluer des rapports d'expériences supposées vraies. Les étudiants étaient randomisés pour recevoir une formation initiale comprenant soit un rationnel positif, c'est-à-dire mettant en avant des études allant dans le même sens que les résultats des expériences présentées, soit un rationnel « négatif » c'est-à-dire mettant en avant des études allant à l'encontre des résultats des expériences présentées. Les étudiants devaient ensuite répondre à des questions permettant d'une part de juger de leur niveau de croyance initiale, et d'autre part portant sur leur perception de la qualité méthodologique des rapports d'expériences supposées vraies. Les résultats de cette expérience ont montré qu'il y avait une concordance entre le jugement scientifique à propos de la qualité méthodologique et les croyances initiales. Les rapports d'expérience montrant des résultats en accord avec les croyances initiales étaient jugés de meilleure qualité que les études qui montraient des résultats discordants[57].

L'excès de confiance accordé par les individus à leur jugement, « overconfidence bias » a surtout été décrit dans les études qui analysent la prise de décision médicale, notamment lors de l'analyse des erreurs diagnostiques[59,60]. En raison d'une mauvaise estimation de

l'étendue et de la justesse de leurs connaissances, certains individus, et plus souvent les experts, ont tendance à sous-estimer le pourcentage d'erreur de leur jugement et à produire des évaluations erronées des risques auxquels ils sont confrontés.

1.2.2.2. Les conflits d'intérêt

De nombreux travaux méthodologiques se sont intéressés d'une part au lien entre les résultats des études et leurs sources de financement, et d'autre part à l'impact sur les lecteurs de la présence ou non de conflit d'intérêt rapporté dans l'étude. C'est ce deuxième point que nous allons détailler ici.

En effet, certains lecteurs sont a priori plus méfiants lors de la lecture d'un rapport d'étude lorsque celle-ci a été financée par l'industrie pharmaceutique. Ils apprécient différemment la qualité méthodologique, l'intérêt et la validité de l'étude selon la présence ou non de conflit d'intérêt.

Plusieurs auteurs ont utilisé des plans expérimentaux pour tenter de répondre à cette question[61,62].

Par exemple, l'étude d'A. Kesselheim publiée en 2012 dans le NEJM, comparait le jugement des lecteurs (médecins internistes) sur la qualité méthodologique à partir de trois résumés d'essais fictifs strictement identiques, hormis la source de financement déclarée. Les lecteurs ont jugé les résumés dont le financement était déclaré industriel comme étant de moins bonne qualité méthodologique que les résumés dont le financement était déclaré provenant d'une organisation gouvernementale (NIH)[63].

Cependant la modification de la perception des résultats liée aux conflits d'intérêts potentiels ou déclarés n'est pas systématiquement retrouvée dans les études explorant ce lien[64]. Ainsi, dans l'essai randomisé contrôlé de C Buffel du Vaure, des médecins généralistes français n'étaient pas influencés lors de la lecture d'un résumé par la présence ou non de source de financement privé ou de conflits d'intérêt déclarés[65].

Les modifications de perception liées aux conflits d'intérêt n'ont, à notre connaissance, jamais été étudiées dans le cadre des revues systématiques et des méta-analyses.

1.2.2.3. Les modes de représentation des résultats

Les résultats de la recherche peuvent être exprimés sous différentes formes : différence absolue, risque relatif, nombre de patients à traiter pour éviter un évènement, etc. Bien que ces différents modes de présentation expriment un même résultat, la perception par le lecteur de l'effet bénéfique du traitement est différente selon le mode de présentation choisi par les auteurs. En effet, la manière dont le résultat d'un ERC est présenté peut avoir une influence d'une part sur la prescription du médecin, et d'autre part sur la promotion du médicament par les décideurs en santé. Par exemple, des études montrent que les médecins, comme les patients, perçoivent comme étant plus bénéfiques les traitements dont les résultats sont présentés sous forme de réduction du risque relatif par rapport à une présentation sous forme de risque absolu[66]. En effet, lorsque le résultat est présenté sous forme de risque relatif le nombre est généralement grand et donc plus marquant que lorsque qu'il est présenté sous forme de risque absolu surtout lorsque le risque absolu est petit. L'utilisation seule du risque relatif peut donc être trompeuse[67]. Par exemple, dans une population où le risque de décès d'une femme par cancer du sein est de 4 pour 1000, et si la mammographie diminue le

risque de décès par cancer du sein de 25%, ce résultat peut être présenté de différentes manières : 1) réduction du risque absolu de décès de 1 pour 1000 ou de 0.001% ; 2) réduction du risque relatif de décès de 25%; 3) parmi 1000 femmes qui ne vont pas participer au dépistage par mammographie, 4 vont décéder d'un cancer du sein et parmi 1000 femmes qui vont participer au dépistage par mammographie, 3 vont décéder d'un cancer du sein et un décès va être évité[67].

La représentation graphique du risque (ou du bénéfice), comme par exemple le choix de la classe de référence, peut aussi avoir un impact important sur la perception du bénéfice. Par exemple la figure 2 ci-dessous représente l'effet d'un traitement par aspirine ou par warfarine chez les patients atteints de fibrillation atriale. Sur la figure de gauche, l'effet du traitement est représenté sous forme de réduction du risque relatif du nombre d'événements et donc relativement à la classe de patients n'ayant pas reçu le traitement. A droite, l'effet traitement est représenté sous forme de réduction du risque absolu, c'est-à-dire relativement à une classe de 100 patients n'ayant pas eu l'évènement.

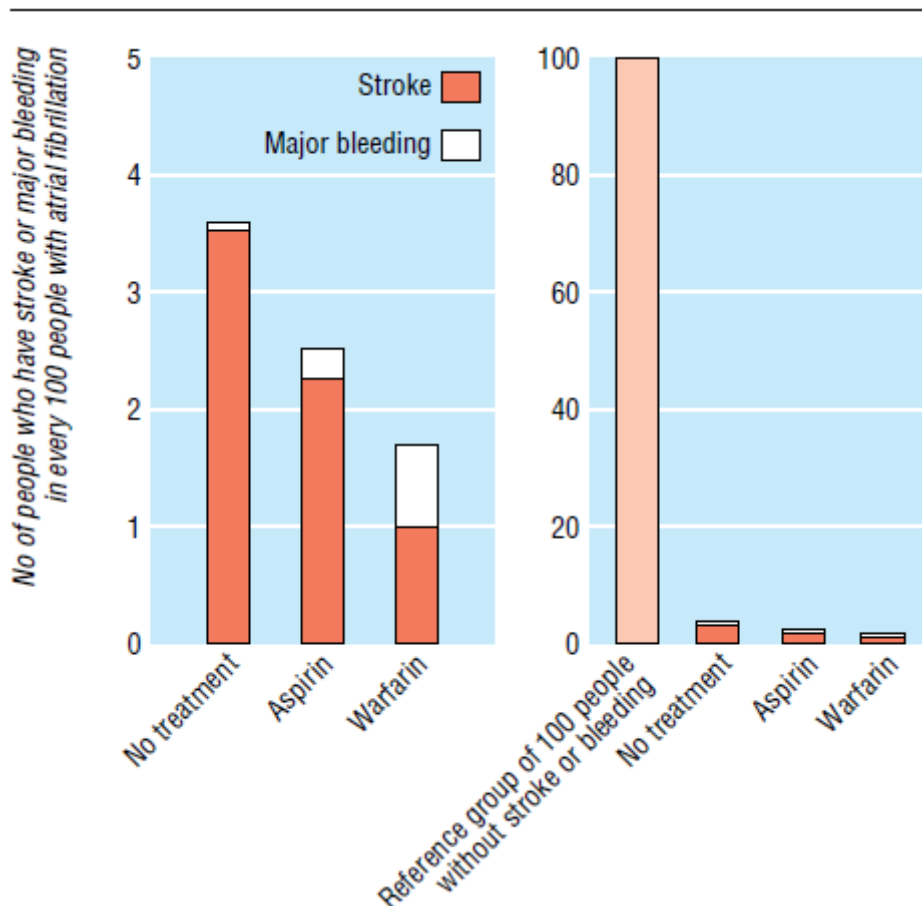


Figure 2 : Différentes représentations du même effet traitement.

Visuellement le bénéfice du traitement semble plus important sur la figure de gauche, c'est-à-dire lorsque la classe de référence des patients sans AVC ou saignement n'est pas représentée.

La perception liée au mode de présentation des résultats a été étudiée par de nombreux auteurs et concerne toutes les catégories de lecteurs quel que soit leur niveau de connaissance médicale ou scientifique : médecins, patients et décideurs en santé sont concernés[68–71].

Ce phénomène a surtout été étudié pour la présentation des résultats des essais randomisés contrôlés. Peu d'auteurs se sont intéressés à ce phénomène dans le cadre des revues systématiques et des méta-analyses. Dans une étude de P. Raina publiée en 2005,

l'interprétation des résultats des méta-analyses varie selon 1) leur mode de présentation, 2) la sévérité de la maladie, 3) la taille de l'effet et 4) la consistance statistique (hétérogénéité). Dans cette étude, les médecins étaient plus confiants dans l'effet bénéfique du traitement quand la taille d'effet était grande par rapport à une taille d'effet plus petite et lorsque la consistance statistique (hétérogénéité) entre les essais inclus dans la méta-analyse était faible[72].

Les discordances d'interprétation des résultats numériques peuvent être liées à des connaissances théoriques ou des expériences différentes entre les auteurs et les lecteurs.

1.2.3. Transparence

L'interprétation des résultats de la recherche est également influencée par la disponibilité des informations accessibles aux lecteurs, c'est-à-dire la transparence de la recherche. Le développement des recommandations pour rapporter les études appelées *reporting guidelines*, ainsi que la politique d'enregistrement des essais ont marqué un tournant dans l'amélioration de la transparence de la recherche.

1.2.3.1. Les *reporting guidelines*

L'interprétation juste des résultats à la lecture d'un article suppose que ceux-ci soient présentés de façon adéquate, c'est-à-dire de manière claire, complète et transparente. Le lecteur peut ainsi juger l'intérêt et la qualité d'un essai et les professionnels de santé peuvent transposer les résultats de la recherche à la pratique clinique.

Aussi des méthodologistes ont établi des recommandations pour l'écriture des rapports d'étude plus communément appelés *reporting guidelines*. Ces recommandations sont regroupées au sein d'un réseau nommé EQUATOR (Enhancing the QUALity and Transparency Of health Research). C'est une initiative internationale qui a pour mission de promouvoir la valeur et la fiabilité de la littérature scientifique médicale par la promotion d'une information précise et transparente des études de recherche[73,74]. Le réseau EQUATOR regroupe des recommandations pour l'écriture de la plupart des types d'études ainsi que leurs mises à jour régulières, comme par exemple les recommandations PRISMA pour l'écriture des revues systématiques[25,75], CONSORT pour les essais randomisés contrôlés[26,76] ou STROBE pour les études observationnelles[77]. Les recommandations concernent le fond et la forme (c'est-à-dire la structure) de l'article et sont présentées à l'aide d'une check-list qui regroupe les items importants de chaque section. Un grand nombre de revues scientifiques exigent aujourd'hui que les auteurs suivent ces recommandations avant de soumettre leur manuscrit. Il existe également des *reporting guidelines* spécifiquement pour l'écriture du résumé d'article (« PRISMA for Abstract »[25], « CONSORT for Abstract »[26,78]. En effet, celui-ci est la principale -voire la seule- source d'informations pour les lecteurs[24]. Il est donc essentiel qu'il rapporte toutes les informations nécessaires et importantes pour juger a minima de la qualité de la recherche présentée.

La description et la discussion des limites d'une étude sont des éléments essentiels qui devraient systématiquement être rapportés lors de la rédaction d'un article scientifique pour pouvoir interpréter correctement les résultats de l'étude[79]. En effet, les limites permettent 1) d'interpréter les résultats à la lumière des biais potentiels de l'étude, 2) de donner un niveau de crédibilité à l'étude et ainsi d'adapter le niveau de preuve, et 3) d'envisager les améliorations à apporter lors des projets de recherche futurs. La description des limites d'une

étude ne devrait pas se contenter de lister l'amplitude et la direction de potentielles erreurs de mesure systématique ou aléatoire. La discussion des limites nécessite une vue d'ensemble de l'étude, de son évaluation et de son interprétation globale et ainsi de donner un sens aux résultats d'une expérimentation. Aussi les recommandations PRISMA pour l'écriture des résumés mentionnent dans la section discussion un item spécifique sur les limites. Cet item définit le contenu des limites à rapporter (Tableau 3)[25].

Les *reporting guidelines*, très largement diffusées, sont publiées simultanément dans de nombreuses revues scientifiques et en accès libre. Aussi la plupart des revues scientifiques s'y réfèrent dans leurs instructions aux auteurs. Il existe des données encourageantes selon lesquelles les *reporting guidelines* améliorent la qualité des articles publiés[29,74,80,81], même si celle-ci peut encore l'être davantage[32–34].

Par contre il n'existe pas aujourd'hui de recommandation pour les auteurs ou les lecteurs sur l'interprétation des résultats issus des recherches scientifiques.

1.2.3.2. Les registres d'essais

Dans le même but de promouvoir la transparence des études scientifiques, le Comité International des Éditeurs de Revues Médicales (ICMJE) promeut l'enregistrement des protocoles des essais dans des registres publics avant le début de l'inclusion du premier patient[82,83]. Cette décision prise en 2004 par les journaux membres de l'ICMJE s'impose à tous les ERC débutants après le 1^{er} juillet 2005. Cette politique éditoriale est soutenue par l'Organisation Mondiale de la Santé ainsi que le gouvernement américain. En effet celui-

ci a rendu obligatoire l'enregistrement des essais pour toute étude dont au moins l'un des investigateurs est américain (FDA Amendement Act 2007)[84]. La plupart des journaux avec un facteur d'impact élevé recommandent ou requièrent désormais cette inscription dans un registre public et libre d'accès.

Le principal registre d'essais cliniques est la base de données ClinicalTrials.gov. Celle-ci a été créée à la suite de la loi sur la modernisation de la Food and Drug Administration en 1997. Cette loi exige que le ministère américain de la Santé et des Services sociaux établisse un registre des essais cliniques pour les études évaluant l'efficacité des médicaments et financés soit par le gouvernement fédéral soit par le secteur privé. Le registre d'essais accessible via le site ClinicalTrials.gov a été mis à la disposition du public en février 2000. En 2010, 330 nouvelles études étaient enregistrées chaque semaine sur le site ClinicalTrials.gov et le registre comprenait près de 80 000 études enregistrées[85].

Sur le même modèle que ClinicalTrials.gov, la base de données PROSPERO a été créée en février 2011[86,87]. PROSPERO est une base de données internationale permettant l'enregistrement prospectif des protocoles de revues systématiques. Elle est hébergée par le *Centre for Reviews and Dissemination* de l'Université de York. PROSPERO a pour objectif de fournir une liste complète des protocoles de revues systématiques pour d'une part éviter les doublons non planifiés et d'autre part améliorer la transparence des revues systématiques en permettant la comparaison de la méthode telle qu'elle est rapportée dans la publication avec celle prévue dans le protocole. En novembre 2014, soit trois ans après sa création, 5 000 revues systématiques étaient enregistrées dans PROSPERO. Contrairement aux essais cliniques, il n'y a pas aujourd'hui de recommandations quant à l'enregistrement des protocoles de revues systématiques.

1.3. Biais de diffusion des résultats

Le biais de diffusion des résultats de la recherche se produit lorsque la diffusion des résultats d'une étude dépend de la direction ou de la force de ses conclusions. Celle-ci est définie comme l'accessibilité ou la possibilité d'être identifié par les lecteurs ou utilisateurs potentiels. Le spectre de diffusion des résultats de la recherche varie entre une totale inaccessibilité (étude jamais publiée ni présentée en congrès ni diffusée par quelque autre moyen) et une accessibilité facilitée (par exemple, une étude publiée en accès libre avec des résultats diffusés sur les réseaux sociaux et dans les médias). La diffusion des résultats dépend de nombreux paramètres comme 1) la présentation en congrès, 2) la publication ou non dans une revue, 3) la diffusion des résultats vers le grand public à travers les réseaux sociaux et les médias généralistes (presse écrite, radio, télévision et Internet)[88]..

Si la diffusion d'un résultat particulier ou d'une étude en général est déterminée par rapport au résultat du test statistique, alors les critères dont le résultat est « statistiquement significatif » du fait du hasard seraient plus fréquemment rapportés et une grande partie de la littérature médicale serait une erreur. En effet, la pratique standard suppose que la différence observée entre les traitements est « statistiquement significative » si la valeur p est inférieure à 0,05 (c'est-à-dire que la probabilité que le résultat obtenu soit due au hasard est inférieure à 5%, en l'absence d'effet vrai). En d'autres termes, si cent tests statistiques sont effectués, cinq associations testées seront « statistiquement significatives » du fait du hasard, même si aucune des associations n'existe vraiment.

Si la littérature biomédicale est une représentation biaisée de ce qui est connu, alors la connaissance médicale basée sur les données de la recherche est elle aussi biaisée. Ainsi l'interprétation des résultats de la recherche dépend non seulement de l'interprétation des

données accessibles mais aussi de la diffusion de données

1.3.1. Le biais de publication des articles

Les biais de publication des articles scientifiques reposent sur le fait que les études qui présentent des résultats « importants » ou statistiquement significatifs sont plus susceptibles d'être publiées que les études qui ne montrent pas de tels résultats. Les résultats « importants » ou statistiquement significatifs sont 1) plus souvent rapportés, 2) plus souvent publiés avec le texte complet accessible, 3) publiés plus rapidement et dans des revues plus prestigieuses, 4) le plus souvent en double exemplaire, 5) en anglais, 6) faciles d'accès, et 7) plus fréquemment cités. Le biais de publication survient principalement avant la présentation des résultats lors de conférences et lors de la soumission des manuscrits aux revues[89–91]. Les biais de publication comprennent 1) les biais liés au délai de publication (un délai de publication plus rapide pour des résultats favorables)[92,93], 2) les biais liés aux publications multiples (multiplication des publications)[94,95], 3) les biais liés à la langue et ceux liés au nombre de citations[96–98] 4) ainsi que les biais liés à l'indexation dans les bases de données. Les biais de publication ont un impact important sur l'estimation globale de l'effet traitement estimé notamment au cours d'une revue systématique avec méta-analyse. Par exemple, dans une étude publiée dans le *New England Journal of Medicine* en 2008, E. Turner[99] a choisi douze antidépresseurs pour lesquels il a comparé les données publiées dans la littérature médicale à celles de la Food and Drug Administration [FDA]. Il a mis en évidence que la majorité des études positives ont été publiées, et que la majorité des études négatives ne l'ont pas été. De plus parmi les 36 études négatives, 11 ont été publiées avec des résultats présentés comme étant positifs. Les auteurs ont également découvert que l'efficacité de ces

médicaments est moins grande que ce que l'on pourrait en conclure en examinant seulement la littérature publiée. L'effet traitement des médicaments est de 11% à 69% plus important dans les données publiées que dans les analyses de la FDA.

Le biais de publication concerne tous les types d'études. Plusieurs auteurs ont quantifié le biais de publication, soit en évaluant le taux de publication des essais présentés en congrès, soit en évaluant le taux de publication des résultats des essais enregistrés sur ClinicalTrials.gov, une fois l'essai terminé.

Dans son étude R. Scherer a montré que globalement moins de 50% des études présentées lors d'un congrès étaient finalement publiées sous forme d'article, soit 63% pour les ERC et 43% pour les autres types d'études[100]. De même, dans l'étude de S. Hopewell, publiée en 2015, 53% des revues systématiques présentées dans neuf congrès internationaux étaient publiées ultérieurement et la nature de la conclusion (en faveur ou non de l'intervention expérimentale) était différente entre le résumé présenté en congrès et la publication dans 24% des cas[36]. L'essor des registres d'essais cliniques avec la possibilité d'enregistrer les résultats des études devrait en théorie permettre de diminuer le biais de publication. Mais les résultats des études sont rarement postés une fois l'essai terminé, dans l'étude de T.A.H. Nguyen publiée en 2013 dans le *Journal of Clinical Oncology*, parmi 209 essais enregistrés dans ClinicalTrials.gov et ayant terminé l'étude, 9% (IC 95%, 7% à 11%) avaient posté les résultats sur le site douze mois après la fin de l'étude, 25% (IC 95%, 21% à 28%) à 24 mois, et 31% (IC 95%, 28% à 35%) à 36 mois [101]. Les registres d'essais ne remplissent que partiellement leur rôle pour diminuer les biais de publication, car bien qu'ils permettent d'identifier les études existantes, les résultats sont rarement rapportés.

1.3.2.La présentation sélective des critères de jugement

Il existe une présentation sélective des critères de jugement lorsque leur publication est orientée par la nature et la direction de leurs résultats. De nombreuses études ont montré l'existence de divergences entre les critères de jugement pré-spécifiés dans le protocole de recherche et les critères de jugement publiés in fine[102–104]. Il est généralement admis que le ou les critères de jugement principaux doivent être définis a priori, c'est-à-dire avant le début de l'étude pour minimiser les biais que pourrait entraîner une sélection post-hoc des résultats. Les divergences entre les critères pré-spécifiés dans les protocoles et les critères publiés peuvent être classées en 5 catégories[105,106] :

- 1) Le critère principal de jugement pré-spécifié est omis dans la publication
- 2) Un nouveau critère principal est introduit dans l'article publié
- 3) La période de mesure des critères de jugement principaux pré-spécifiés enregistrée et publiée diffère
- 4) Le critère de jugement principal pré-spécifié est cité comme un critère de jugement secondaire dans l'article publié
- 5) Un critère de jugement secondaire pré-spécifié est décrit comme principal dans l'article publié.

Malgré une augmentation importante de l'enregistrement des essais dans les registres public, le biais de publication des critères de jugement reste fréquent ; il est de l'ordre de 40% à 60% selon les études[103,107–111]. Le biais de publication de critères de jugement concerne également les critères qui sont rapportés de manière incomplète, c'est-à-dire qui ne peuvent pas être intégrés dans une future méta-analyse. Dans l'étude de A.W. Chan, portant sur 102 protocoles d'ERC et leur publication, 50% des critères d'efficacité et 65% des critères de

sécurité sont rapportés de manière incomplète[112]. Le défaut de publication des critères de sécurité et des effets indésirables en général peut avoir de lourdes conséquences pour les patients. Plusieurs « affaires » hautement médiatisées ont mis en exergue cette problématique, comme par exemple le cas du rofécoxib (Vioxx®). L'étude princeps ayant permis l'autorisation de mise sur le marché de ce médicament a rapporté des résultats tronqués et dont la présentation fausse le jugement du lecteur et de nombreux effets indésirables ne sont pas rapportés[113,114].

Malgré les recommandations de l'ICMJE pour l'enregistrement des essais cliniques, un nombre non négligeable d'ERC publiés après 2005 n'a jamais été enregistré. En effet, au sein de deux cohortes d'ERC, l'une de 327 ERC publiés dans des journaux chirurgicaux[115] et l'autre de 197 ERC concernant la réanimation[110] seuls les deux tiers des essais étaient enregistrés. Aussi l'enregistrement reste un outil passif et il ne peut à lui seul prévenir les biais de publication. L'existence de différences entre le protocole enregistré et la publication finale n'est pas rare[116] et a été décrit dans de nombreux domaines de la médecine [111,112,117].

1.3.3. Les *spin*

Des stratégies utilisées par les auteurs pour « enjoliver » leurs données ont depuis longtemps été décrites. Elles concernent la manière dont sont rapportées les études et ont été récemment regroupées sous le nom de *spin*.

On ne sait pas si ces stratégies sont à l'origine d'une interprétation biaisée des résultats.

1.3.3.1. Définition du concept de *spin*

Le concept de *spin* est nouveau en recherche médicale, il fut employé pour la première fois dans la littérature médicale en 1995 par R. Horton dans son article sur la rhétorique de la recherche : « the spin that authors place on their own work. »[118]. Celui-ci décrit les « astuces » conscientes ou inconscientes qui influencent l'interprétation des résultats : « to interpret a result correctly reviewers, statisticians, editors, and readers should know the conscious and unconscious tricks of authorial rhetoric ».

Nous avons défini les *spin* comme une distorsion de la présentation et de l'interprétation des résultats de la recherche par analogie aux *spin doctors* qui adaptent les messages par une distorsion des faits ou des informations pour les présenter sous un angle favorable et ainsi influencer l'opinion. Dans notre domaine d'intérêt, les *spin* visent à enjoliver les résultats de la recherche (« beautification of data ») et à convaincre le lecteur d'un article que l'effet bénéfique du traitement évalué est plus important que ce que montrent réellement les résultats rapportés dans l'article.

La rhétorique fait partie de la tradition scientifique. La communication scientifique permet à

la description factuelle des résultats d'une étude, mais aussi la transmission un message et d'une conviction. Les auteurs ont une grande latitude dans la rédaction d'un article scientifique et vont utiliser la rhétorique d'une part pour transmettre un savoir, c'est-à-dire le résultat de l'expérience, et d'autre part faire passer un message, c'est à dire l'interprétation de ce résultat. La force de persuasion d'un texte dépend de sa structure passive ou active, de la forme narrative, du temps employé (présent, futur) et des métaphores utilisées[119]. Par exemple, lorsqu'un auteur décrit un résultat négatif en ces termes : « The treatment effect approached, but did not achieve conventional statistical significance »[120] cela signifie que le résultat est statistiquement non significatif mais semble bien plus positif ou optimiste que l'énoncé factuel « our results were not statistically significant ».

1.3.3.2. Les *spin* dans les essais randomisés contrôlés

Les premiers travaux de recherche sur les *spin* ont porté sur les articles rapportant les résultats d'ERC ; ce type d'étude suit une méthodologie très standardisée. Dans cette catégorie d'article, les *spin* sont les modes de présentation ou d'interprétation des résultats par les auteurs qui mettent en avant un éventuel effet bénéfique de l'intervention expérimentale[34]. Dans cette étude I. Boutron a identifié la nature des *spin* dans les ERC dont le résultat du critère principal de jugement est statistiquement non significatif et a évalué leur fréquence d'utilisation. Les *spin* ont été identifiés à partir d'une cohorte représentative d'ERC dont le critère principal de jugement est statistiquement non significatif et indexé dans PubMed. La classification des *spin* a ensuite été développée à partir des *spin* retrouvés dans les ERC et enrichie par une revue de littérature sur ce thème ainsi qu'une enquête auprès des membres du groupe méthode statistique de la Collaboration Cochrane. Dans la classification proposée par

I. Boutron les *spin* sont regroupés selon la section de l'article où ils sont retrouvés (Tableau 2). Cette classification a par la suite été reprise par différents auteurs pour évaluer la prévalence des *spin* dans leur domaine de spécialité comme par exemple S. Lockyer[121] qui a décrit les *spin* dans les ERC étudiant les soins de cicatrisation.

Les auteurs ont également défini un niveau de sévérité des *spin* dans la conclusion : sévère, modéré ou mineur. La sévérité des *spin* dans la conclusion dépend de 3 critères: 1) la présence ou non d'incertitude dans la construction sémantique de la conclusion, 2) la présence ou non de recommandations pour effectuer de nouvelles études, et 3) la reconnaissance ou non du caractère non statistiquement significatif du ou des critères principaux de jugement.

Parmi les types de *spin* décrits dans les ERC dont le critère principal de jugement est statistiquement non significatif, nous pouvons citer par exemple: 1) le fait de conclure en s'appuyant uniquement sur les critères de jugement dont les résultats sont statistiquement significatifs, 2) ne pas mentionner des effets indésirables importants, 3) présenter sélectivement les critères statistiquement significatifs dans le résumé ou 4) rapporter un critère statistiquement non significatif comme s'il s'agissait d'un essai d'équivalence.

Pour illustrer les *spin*, nous prendrons pour exemple l'article de Lisa de la Fuentes publié en 2009 dans *Journal of American of College of Cardiology*[122]. Cet essai randomisé contrôlé avait pour objectif d'évaluer l'effet de deux types de régimes (régime pauvre en carbohydrates et régime pauvre en graisses) sur les fonctions et structures cardiovasculaires chez les obèses. Dans la section « résultats » de l'article les auteurs rapportent qu'il n'y a aucune différence entre les deux groupes en termes de perte de poids et pour tout autre critère principal : « *There were no differences between the 2 diet groups in terms of weight loss or any of the primary*

end points, namely diet-induced changes in cardiac structure/function and CIMT; therefore, the data for the 2 groups were combined ». Cependant les auteurs concluent dans la section conclusion du résumé : « ***Diet-induced moderate weight loss in obese subjects is associated with beneficial changes in cardiovascular structure and function. Subsequent weight regain is associated with partial loss of these beneficial effects*** ». Dans cet exemple, les auteurs 1) rapportent dans le résumé uniquement les résultats statistiquement significatifs (alors même qu'il s'agit d'analyses post-hoc), 2) ne mentionnent pas que les critères de jugement principaux sont non statistiquement significatifs et enfin 3) concluent en mettant en avant le résultat d'une comparaison inadéquate, une analyse intra-groupe (avant-après) qui associe les patients des deux bras en un seul groupe.

Cet autre exemple, illustre les *spin* liés à une interprétation inappropriée des résultats, comme par exemple le fait de rapporter un critère non statistiquement significatif comme si l'essai était un essai d'équivalence. Cette étude de W. Nienhuis publiée en 2010, est un ERC qui compare l'efficacité de deux traitements médicamenteux sur les infections à *mycobacterium ulcerans*. Il n'y a pas de différence statistiquement significative entre les deux groupes pour le critère de jugement principal - la guérison à un an[123]. Cependant les auteurs concluent que les deux traitements sont efficaces et ont la même efficacité : « *Interpretation :Antimycobacterial treatment for M ulcerans infection is effective in early, limited disease. 4 weeks of streptomycin and rifampicin followed by 4 weeks of rifampicin and clarithromycin has similar efficacy to 8 weeks of streptomycin and rifampicin* ». Cette conclusion peut porter à confusion car l'absence de différence retrouvée peut être liée soit à une absence vraie de différence soit à une différence non retrouvée dans cette étude, par manque de puissance. Aussi l'absence de différence retrouvée entre les deux groupes n'est pas synonyme d'efficacité, en effet il n'y a pas de comparaison avec un placebo et on peut

supposer que les deux traitements sont pareillement inefficaces.

Tableau 2: Classification des *spin* dans les ERC

<p><i>Spin</i> dans la section résultat</p> <ul style="list-style-type: none"> - Focaliser sur une comparaison intra groupe statistiquement significative - Focaliser sur un critère secondaire statistiquement significatif - Focaliser sur une analyse en sous-groupe statistiquement significative - Focaliser sur une analyse sur une population modifiée statistiquement significative (par exemple : analyse per protocole) - Focaliser sur une comparaison intra ou inter groupe d'un critère secondaire statistiquement significatif - Autre type de <i>spin</i> dans la section résultats
<p><i>Spin</i> dans la section discussion</p> <ul style="list-style-type: none"> - Rapporter un critère non statistiquement significatif comme si l'essai était un essai d'équivalence - Focaliser sur un critère secondaire statistiquement significatif - Focaliser sur une analyse en sous-groupe statistiquement significative - Focaliser sur une population d'analyse modifiée - Focaliser sur une amélioration inter groupe globale - Omettre des effets indésirables - Autre types de <i>spin</i> dans la discussion
<p><i>Spin</i> dans la section conclusion</p> <ul style="list-style-type: none"> - Focaliser uniquement sur les résultats statistiquement significatifs - Affirmer l'équivalence des traitements pour des résultats statistiquement non significatifs - Affirmer l'efficacité sans considération pour le critère principal statistiquement non significatif - Présenter les résultats statistiquement non significatifs du critère primaire, tout en soulignant l'effet bénéfique du traitement - Présenter les résultats statistiquement non significatifs du critère primaire, mais mettre en avant les autres résultats statistiquement significatifs - Omission des effets indésirables - Focaliser sur une évaluation intra groupe - Recommandation pour utiliser le traitement - Focaliser sur un autre objectif - Comparaison avec un groupe placebo venant d'une autre étude - Rapporter un résultat statistiquement non significatif d'un sous groupe comme étant bénéfique

1.3.3.3. Les *spin* dans les études diagnostiques

L'étude d'E.A. Ochodo s'est, quant à elle, intéressée aux *spin* dans les études sur la précision diagnostique[124]. Les auteurs ont élaboré une classification en deux parties : 1) les *spin* aboutissant à une surinterprétation des résultats de l'étude (« actual overinterpretation in diagnostic accuracy studies »), qui contient six types de *spin* différents, et 2) les *spin* pouvant potentiellement aboutir à une surinterprétation des résultats de l'étude (« potential overinterpretation in diagnostic accuracy studies ») qui contient huit *spin*. Cette classification est présentée dans le tableau 3.

Les auteurs définissent la surinterprétation des résultats de l'étude lorsque les résultats semblent plus favorables que ce que justifient les résultats. Parmi ce type de *spin* on retrouve par exemple, la présentation sélective dans le résumé des critères dont les résultats sont les plus en faveur de l'effet bénéfique de l'intervention ou lorsque la conclusion ne reflète pas l'objectif de l'étude. Par exemple dans cette étude [125] qui avait pour objectif d'évaluer la sensibilité et la spécificité du test EV71 IgM pour le diagnostic précoce de la maladie pieds-main-bouche, les auteurs concluent que la détection des IgM anti-EV71 par ELISA permet un diagnostic facile, rapide et fiable. La conclusion des auteurs ne porte pas sur la précision diagnostique mais sur d'autres mesures de performance diagnostique.

Les *spin* pouvant potentiellement aboutir à une surinterprétation des résultats de l'étude sont définis comme l'absence de description des éléments importants de la méthodologie, comme les hypothèses sous-jacentes à l'étude, le calcul du nombre de sujets à inclure, les seuils de positivité des tests, ou l'emploi de tests statistiques inappropriés.

Tableau 3: Classification des *spin* dans les études de précision diagnostique

<p>Stratégies de <i>spin</i> aboutissant à une surinterprétation des résultats de l'étude</p> <ul style="list-style-type: none"> - Résumé trop optimiste - Conclusion plus « forte » dans le résumé - Présentation sélective des résultats dans le résumé - Conclusion basée sur des sous groupes choisis - Disparité entre l'objectif et la conclusion - Articles avec au moins une stratégie de surinterprétation
<p><i>spin</i> pouvant potentiellement aboutir à une surinterprétation des résultats de l'étude</p> <ul style="list-style-type: none"> - Calcul du nombre de sujets nécessaires non rapporté - Hypothèse de test non pré-spécifiée - Intervalles de confiance des mesures de la précision non rapportés - Rôle des tests non rapporté ou mal défini - Groupes pour analyses en sous-groupe non pré-spécifiés dans la section méthode - Seuils de positivité des tests continus non rapportés - Utilisation de tests statistiques inappropriés - Articles avec au moins une stratégie de surinterprétation potentielle

1.3.3.4. Les *spin* dans les études observationnelles et les revues systématiques

Aucune étude n'a à notre connaissance décrit de manière systématique les *spin* dans les études observationnelles ou dans les revues systématiques avec ou sans méta-analyses.

Cependant différents types de *spin* ont été décrits sans toutefois proposer de classification.

Les articles de S. Cofield[126], de A.W. Brown[127] ou de N. Menachemi[128] ont décrit plus particulièrement l'utilisation inappropriée de termes de causalité dans des études observationnelles. En effet les études observationnelles ne permettent pas de montrer une inférence causale mais uniquement une association.

Afin d'illustrer ce type de *spin*, nous prendrons pour exemple l'étude de R. Kenter publiée en 2013 dans *BMC psychiatry*[129]. Dans cette étude observationnelle, parmi 104 patients sélectionnés présentant des symptômes de dépression ou d'anxiété sévère, 55 acceptent de participer à l'étude. L'intervention étudiée consiste en un traitement guidé « en ligne » pendant 5 semaines. Les résultats de l'étude montrent qu'après 5 semaines de traitement les symptômes de dépression et d'anxiété ont diminué par rapport aux patients qui ont refusé de participer à l'étude, cette différence n'est pas retrouvée à 12 semaines. Les auteurs concluent dans le résumé: « [...]on line treatment increases speed of recovery and can therefore be offered as a first step of treatment in mental healthcare ». Les auteurs concluent à l'efficacité du traitement alors que d'une part il s'agit d'une étude observationnelle et que d'autre part il n'y a pas d'association retrouvée entre l'intervention et les symptômes de dépression et d'anxiété à 12 semaines.

Dans un autre article publié en 2012, P. Cummings a décrit les *spin* qui consistent à enjoliver ou à magnifier les résultats de la recherche[130]. Ce type de *spin* est également appelé *spin* linguistique ; il s'agit par exemple de l'utilisation de mots pour convaincre au-delà de ce qui a été réellement démontré, « nos résultats *démontrent clairement* que.... » ou « notre étude montre l'association *robuste* entre » ou le fait de se targuer d'avoir été la *première* étude, le *plus grand* échantillon, la *plus forte* preuve de [...]. En effet, les auteurs ont une grande latitude dans la rédaction d'un article scientifique et vont utiliser la rhétorique d'une part pour

transmettre un savoir, c'est-à-dire le résultat de l'expérience et d'autre part faire passer un message c'est à dire l'interprétation de ce résultat. La force de persuasion d'un texte dépend de sa structure passive ou active, de la forme narrative, du temps employé (présent, futur) et des métaphores utilisées[119]. Par exemple, lorsqu'un auteur décrit un résultat négatif en ces termes : « The treatment effect approached, but did not achieve conventional statistical significance »[120] cela signifie que le résultat est statistiquement non significatif mais semble bien plus positif ou optimiste que l'énoncé factuel "our results were not statistically significant".

1.3.3.5. Prévalence des *spin*

La classification des *spin* décrite par I. Boutron a permis pour la première fois d'évaluer leur prévalence. Dans cette cohorte de 72 ERC, les auteurs retrouvaient un *spin* dans 18% des titres, et au moins un type de *spin* dans 37,5% des cas pour la section résultats du résumé et 29,2% pour la section résultats du texte intégral ; 58,3% pour la conclusion du résumé et 50% pour la conclusion du texte intégrale. Les *spin* les plus fréquents étaient : 1) focaliser l'attention sur une comparaison intra groupe statistiquement significative dans 14% des ERC, 2) affirmer l'équivalence des traitements quand les résultats sont statistiquement non significatifs dans 14% des ERC et 3) focaliser l'attention sur une amélioration intergroupe dans 12,5 % des ERC. De plus, parmi les ERC présentant des *spin* dans la conclusion du résumé, un tiers présentaient un *spin* sévère, c'est-à-dire considérés comme étant particulièrement graves et influents[34]. Cette classification a été utilisée ensuite par des chercheurs de différentes disciplines pour évaluer les *spin* dans les essais publiés dans leur domaine. J Wilson a décrit les *spin* utilisés pour minimiser les risques dans les essais évaluant

le bénéfice d'un défibrillateur implantable en prévention primaire des troubles du rythme. Wilson a décrit la présence d'au moins un *spin* dans les dix études incluses dans cette étude[131]. Dans la cohorte d'E.A. Ochodo qui comprenait 126 ERC diagnostiques, 31% (n=39; IC 95% [23-39]) contenaient au moins une stratégie de *spin* et dans un sous-groupe d'étude d'imagerie, 30% (n=16, IC 95% [17-43]) contenaient un *spin* sévère et 100% un *spin* « potentiel ».

Ces différents travaux ont confirmé que des *spin* sont fréquemment retrouvés dans les publications scientifiques. Aussi la section résumé, et plus particulièrement la conclusion du résumé, est la section de l'article où les *spin* sont les plus fréquents.

1.4. Justification de la démarche de thèse

Les *spin* sont un phénomène ancien mais décrit de manière systématique depuis peu. Nous faisons l'hypothèse que les *spin* sont présents, non seulement dans les articles scientifiques, mais également dans les supports de diffusion des résultats scientifiques comme les communiqués de presse ou les articles de presse. Ainsi, les *spin* pourraient potentiellement biaiser l'interprétation des résultats scientifiques. Nous avons réalisé une étude de cohorte pour 1) évaluer la présence des *spin* dans les communiqués et les articles de presse relatant les résultats d'ERC et 2) étudier l'impact de ces *spin* sur l'interprétation des résultats par les lecteurs.

Aussi, plusieurs travaux se sont intéressés aux différents éléments pouvant modifier l'interprétation des résultats de la recherche. Les éditeurs de revues médicales, chercheurs et méthodologistes se sont associés pour élaborer des recommandations pour rapporter les études afin d'améliorer la qualité et la transparence de la recherche. Certaines revues médicales, comme *Annals of Internal Medicine*, ont pris l'initiative d'imposer aux auteurs l'écriture d'une section « limites » dans le résumé pour pouvoir soumettre un article pour publication. Cependant l'impact de cette recommandation n'a à notre connaissance jamais été évalué. Dans un deuxième travail nous avons réalisé un essai randomisé contrôlé étudiant l'impact de l'ajout d'une section limite dans les résumés des revues systématiques sur l'interprétation des résultats par les lecteurs.

Ensuite, comme nous l'avons vu, les revues systématiques et les méta-analyses sont des outils essentiels pour évaluer les interventions de santé et pour identifier les futures recherches à mener. Aussi, afin que les revues systématiques puissent jouer pleinement leur rôle celles-ci

doivent d'une part être réalisées de manière rigoureuse et d'autre part présenter des résultats correctement interprétés. Or l'interprétation des résultats est complexe et les conséquences de cette interprétation sont importantes pour les patients. Il n'existe pas à notre connaissance de classification des *spin* dans les revues systématiques et les méta-analyses. Dans une troisième étude nous avons 1) répertorié les différents *spin* dans le résumé et dans le texte intégral des revues systématiques et des méta-analyses et 2) réalisé une enquête pour classer les *spin* des résumés selon leur probabilité à fausser l'interprétation des lecteurs.

Pour chaque travail, nous résumerons les méthodes ainsi que les principaux résultats avant de présenter le tiré-à-part de la publication ou pour le dernier travail, une version préliminaire de l'article. Finalement, le chapitre 5 de cette thèse sera consacré à la discussion de nos résultats et les perspectives de nos travaux de recherche.

2. Interprétation des résultats des essais randomisés contrôlés dans les communiqués de presse et les articles de presse : une étude de cohorte.

2.1. Justification de l'étude

Les médias de masse jouent un rôle important dans la diffusion des résultats de la recherche clinique. Chaque jour des articles sont publiés dans les journaux, les magazines ou sur internet avec pour objectif de transmettre l'information scientifique au plus grand nombre. Ces articles sont écrits par des journalistes qui souvent s'appuient sur les communiqués de presse[46,132].

La plupart des études sur les communiqués et les articles de presse se sont intéressées à la qualité des informations données (précision des résultats, limites de l'étude, extrapolation des résultats de l'animal à l'homme...)[43,49,133] mais aucune n'a étudié spécifiquement la présence de *spin*, ni l'impact des *spin* sur l'interprétation des résultats par les lecteurs.

2.2. Objectif

Les objectifs de ce premier travail étaient 1) d'évaluer la présence des *spin* dans les communiqués de presse et dans les articles de presse qui relatent les résultats des essais randomisés contrôlés, et 2) d'étudier l'impact de la présence des *spin* sur l'interprétation des résultats par les lecteurs.

2.3. Méthode

Nous avons réalisé une étude de cohorte pour répondre à cette question.

Dans un premier temps, nous avons constitué une cohorte de communiqués de presse relatant

les résultats d'essais randomisés contrôlés (ERC). Nous avons inclus tous les communiqués de presse présentant les résultats d'essais prospectifs randomisés contrôlés à deux bras parallèles évaluant des interventions de soins de santé chez l'Homme, publiés et indexés dans la base de données « EurekaAlert! » entre le 1^{er} décembre 2009 et le 31 Mars 2010.

Nous avons recherché la publication rapportant les résultats de l'ERC décrits dans le communiqué de presse, puis nous avons recherché l'existence d'article de presse décrivant les résultats de ce même essai dans la base de données Lexis-Nexis.

Après avoir extrait les caractéristiques générales de l'essai (mode de financement, nombre de patients inclus, type de traitement expérimental, etc) nous avons recherché de manière systématique la présence de *spin* 1) dans le communiqué de presse, 2) dans l'article de presse et 3) dans la conclusion du résumé de la publication scientifique. Nous avons recherché les *spin* uniquement dans la conclusion du résumé car 1) le résumé est la partie de l'article accessible gratuitement, 2) les lecteurs forment leur opinion le plus souvent à partir des informations données la conclusion du résumé[78] et 3) la conclusion du résumé est la section de l'article où les *spin* sont les plus fréquents[34].

Nous avons utilisé la classification des *spin* décrite dans l'encadré ci-dessous.

Tableau 4: Classification des *spin*

- Omission du critère principal de jugement non statistiquement significatif
- Résultat non statistiquement significatif rapporté comme s'il était statistiquement significatif
- Revendiquer à tort l'équivalence des traitements alors que le résultat des tests ne montre de différence statistiquement significative
- Accent mis sur un critère secondaire statistiquement significatif
- Accent mis sur une comparaison inappropriée de sous-groupes
- Accent mis sur une comparaison intra-groupe
- Résultat négatif rapporté avec un "spin" linguistique
- Ignorer les effets indésirables
- Bonne tolérance d'un traitement revendiqué à tort
- Extrapoler de manière inappropriée les résultats d'un essai

Dans un deuxième temps, les résultats des essais étaient interprétés indépendamment selon les trois sources différentes: 1) la publication scientifique, 2) le communiqué de presse et 3) l'article de presse. Pour chaque source, deux chercheurs ont évalué indépendamment les résultats en prenant en compte les critères de jugement principaux et secondaires ainsi que les données de sécurité. Pour interpréter les résultats des essais nous avons utilisé une échelle allant de 1 à 5 : (1) le patient devrait définitivement prendre le traitement, (2) le patient devrait probablement prendre le traitement, (3) le patient devrait décider pour lui-même (neutre), (4) le patient ne devrait probablement pas prendre le traitement et (5) le patient ne

devrait définitivement pas prendre le traitement.

Nous avons défini comme biaisée, une interprétation du communiqué de presse ou de l'article de presse qui surestime le bénéfice du traitement expérimental. C'est-à-dire, si les résultats du RCT sont interprétés comme neutre à la lecture de l'article, une interprétation du communiqué de presse en faveur du traitement expérimental est une interprétation biaisée des résultats. Et si les résultats du RCT sont interprétés comme en défaveur du traitement expérimental, il y a un biais d'interprétation quand le communiqué de presse est interprété comme neutre ou en faveur du traitement expérimental.

Nous avons planifié des analyses univariées et multivariées pour identifier les facteurs associés à la présence de *spin* ainsi que les facteurs associés à un biais d'interprétation des résultats de l'essai dans les communiqués et les articles de presse.

2.4. Résultats

La cohorte d'analyse comportait 70 couples de communiqués de presse et d'articles scientifiques publiés rapportant les résultats de 70 ERC ainsi que 41 articles de presse relatant l'un des ERC de la cohorte.

Dans notre cohorte, 40% (n=28) des articles scientifiques publiés et 47% (n=33) des communiqués de presse comportaient au moins un *spin*. Les *spin* les plus fréquemment retrouvées étaient 1) l'omission d'un des critères principaux de jugement non statistiquement significatifs, 2) une extrapolation inappropriée des résultats, 3) l'accent mis sur une comparaison intra-groupe ou 4) la revendication à tort de l'équivalence des traitements quand

les tests ne montrent pas de différence statistiquement significative. Le seul facteur associé à la présence de *spin* dans les communiqués de presse – en analyse multivariée - était la présence de *spin* dans la conclusion du résumé de l'article scientifique (RR=5.6, IC 95% [2.8-11.0], $p<0.001$).

L'interprétation des communiqués de presse par les lecteurs était biaisée (c'est-à-dire une interprétation différente de celle de l'article scientifique) pour 22 (31%) d'entre eux. Le traitement expérimental était interprété comme 1) bénéfique pour le patient pour 38 articles (54%) et 55 communiqués de presse (79%), 2) neutre pour le patient pour 18 articles (26%) et deux communiqués de presse (3%) et 3) non bénéfique pour le patient pour 14 articles (20%) et 13 communiqués de presse (18%).

Dans 19 cas (86%) les lecteurs surestimaient l'effet bénéfique du traitement expérimental à la lecture du communiqué de presse par rapport à la lecture de l'article scientifique. Dans trois cas (14%) les lecteurs ont sous-estimé l'effet du traitement à la lecture du communiqué de presse. Les facteurs associés - en analyse multivariée - à un biais d'interprétation des résultats dans les communiqués de presse étaient : 1) la publication de l'essai dans un journal spécialisé (6% versus 45%, $p<0.001$), 2) les essais avec un petit effectif, c'est-à-dire $n<112$ (46% versus 9%, $p<0.001$) et 3) la présence de *spin* dans le communiqué de presse (48% versus 8%, $p<0.001$).

Parmi les 41 essais pour lesquels nous avons retrouvé un article de presse associé au communiqué de presse et à l'article scientifique publié, au moins un *spin* était identifié dans 17 (41%) résumés d'article, 19 (46%) communiqués de presse et 21 (51%) articles de presse. Il y avait au moins un *spin* dans tous les communiqués de presse et dans tous les articles de

presse qui relaient un ERC dont l'article scientifique publié contenait également un *spin*. Les facteurs associés à la présence de *spin* dans les articles de presse étaient la publication de l'essai dans une revue spécialisée (67% versus 35%, $p=0,04$), un petit effectif (68% versus 32%, $p=0,02$), la présence de *spin* dans le résumé de l'article scientifique (100% versus 13%, $p<0,001$) et la présence de *spin* dans le communiqué de presse (100% versus 13%, $p<0,001$). Aussi les lecteurs surestimaient l'effet bénéfique du traitement à la lecture de l'article par rapport à l'article scientifique dans 10 cas (24%). Le biais d'interprétation des résultats de l'essai à la lecture de l'article de presse était associé à 1) un petit effectif (41% vs 5%, $p=0,01$) et 2) la présence de *spin* dans l'article de presse (43% versus 5%, $p=0,009$).

2.5. Discussion

Nous avons montré dans cette étude que la présence de *spin* dans les communiqués et articles de presse est corrélée à la présence de *spin* dans les résumés d'article scientifique publié. De plus, la présence de *spin* dans les communiqués et les articles de presse est associée à une interprétation biaisée des résultats de l'ERC par les lecteurs et le biais d'interprétation est majoritairement dans le sens d'une surestimation de l'effet bénéfique du traitement à l'étude. Nous faisons l'hypothèse que l'interprétation biaisée des résultats des ERC dans les communiqués et articles de presse est liée à la présence de *spin* et que ces *spin* proviennent des publications scientifiques.

Les communiqués de presse jouent un rôle majeur dans la communication et la diffusion des résultats scientifiques[37]. Notre étude confirme que la qualité des communiqués de presse est médiocre, les informations transmises sont souvent tronquées voire erronées. La présence de

spin devient problématique si elle modifie l'interprétation des résultats par les lecteurs. Notre étude fait le lien entre une présentation biaisée des résultats dans les publications scientifiques et une interprétation biaisée de ces résultats dans les médias. Cela met en évidence la nécessité d'une relecture de qualité avant publication afin de détecter la présence de *spin* et ainsi éviter leur diffusion.

Notre étude comporte plusieurs limites. Premièrement, notre cohorte n'est pas représentative de tous les communiqués de presse d'ERC. En effet notre cohorte ne contient que des ERC avec un communiqué de presse indexé dans EurekaAlert! Cependant la base de données EurekaAlert! est une des plus importantes bases de données de communiqués de presse scientifiques utilisée par les journalistes. Deuxièmement, les ERC ne représentent qu'une petite partie de la littérature scientifique médicale et nos résultats ne sont pas transposables aux autres types d'essais. Troisièmement, nous avons cherché les *spin* uniquement dans la conclusion du résumé des ERC et non dans le texte intégral de la publication. En conséquence nous ne pouvons pas affirmer que les *spin* retrouvés dans les communiqués et les articles de presse étaient les mêmes que les *spin* présents dans la publication scientifique. Nous avons choisi de nous concentrer sur la conclusion du résumé car il s'agit de la section de l'article disponible le plus aisément et gratuitement. De plus de nombreux lecteurs utilisent la conclusion du résumé comme message à retenir sans analyse critique de l'article complet. Dernièrement, l'interprétation des résultats d'un essai est toujours au moins en partie subjective. Pour pallier à ce problème toutes les évaluations ont été réalisées par deux personnes indépendantes et les discordances ont été résolues par consensus.

Pour conclure, la présence de *spin* dans les communiqués et articles de presse est corrélée à la présence de *spin* dans les publications scientifiques et est associée à une interprétation biaisée

de résultats de la recherche. Par conséquent, les éditeurs et les relecteurs d'articles scientifiques ont un rôle important à jouer dans la diffusion des résultats scientifiques et devraient être particulièrement attentifs à la présentation et à l'interprétation des résultats par les auteurs dans leurs articles.

2.6. Article

Yavchitz Amélie, Isabelle Boutron, Aida Bafeta, Ibrahim Marroun, Pierre Charles, Jean Mantz, and Philippe Ravaud. 2012. "Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study." Edited by Lisa A. Bero. *PLoS Medicine* 9 (9): e1001308. doi:10.1371/journal.pmed.1001308.

Ce travail a fait l'objet d'un éditorial publié dans *Plos Medicine* et de plusieurs articles de presse. De plus cet article a été l'article le plus tweeté de Plos Medicine selon ses éditeurs et fait partie du top 1% des articles scientifiques les plus tweetés selon ALTMETRIC.

Score ALTMETRIC : 412

Index de citation (ISI web of science): 26

Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study

Amélie Yavchitz^{1,2,3}, Isabelle Boutron^{1,2,3*}, Aida Bafeta^{1,2,3}, Ibrahim Marroun⁴, Pierre Charles⁴, Jean Mantz⁵, Philippe Ravaud^{1,2,3}

1 INSERM, U738, Paris, France, **2** Centre d'Épidémiologie Clinique, AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Hôtel Dieu, Paris, France, **3** Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris, France, **4** Department of Internal Medicine, Hôpital Foch, Suresnes, France, **5** Department of Anesthesiology and Critical Care, Beaujon University Hospital, Clichy, France

Abstract

Background: Previous studies indicate that in published reports, trial results can be distorted by the use of "spin" (specific reporting strategies, intentional or unintentional, emphasizing the beneficial effect of the experimental treatment). We aimed to (1) evaluate the presence of "spin" in press releases and associated media coverage; and (2) evaluate whether findings of randomized controlled trials (RCTs) based on press releases and media coverage are misinterpreted.

Methods and Findings: We systematically searched for all press releases indexed in the EurekAlert! database between December 2009 and March 2010. Of the 498 press releases retrieved and screened, we included press releases for all two-arm, parallel-group RCTs ($n=70$). We obtained a copy of the scientific article to which the press release related and we systematically searched for related news items using Lexis Nexis. "Spin," defined as specific reporting strategies (intentional or unintentional) emphasizing the beneficial effect of the experimental treatment, was identified in 28 (40%) scientific article abstract conclusions and in 33 (47%) press releases. From bivariate and multivariable analysis assessing the journal type, funding source, sample size, type of treatment (drug or other), results of the primary outcomes (all nonstatistically significant versus other), author of the press release, and the presence of "spin" in the abstract conclusion, the only factor associated with "spin" in the press release was "spin" in the article abstract conclusions (relative risk [RR] 5.6, [95% CI 2.8–11.1], $p<0.001$). Findings of RCTs based on press releases were overestimated for 19 (27%) reports. News items were identified for 41 RCTs; 21 (51%) were reported with "spin," mainly the same type of "spin" as those identified in the press release and article abstract conclusion. Findings of RCTs based on the news item was overestimated for ten (24%) reports.

Conclusion: "Spin" was identified in about half of press releases and media coverage. In multivariable analysis, the main factor associated with "spin" in press releases was the presence of "spin" in the article abstract conclusion.

Please see later in the article for the Editors' Summary.

Citation: Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, et al. (2012) Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. *PLoS Med* 9(9): e1001308. doi:10.1371/journal.pmed.1001308

Academic Editor: Lisa A. Bero, University of California, San Francisco, United States of America

Received: November 16, 2011; **Accepted:** August 1, 2012; **Published:** September 11, 2012

Copyright: © 2012 Yavchitz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: No direct funding was received for this study. The authors were personally salaried by their institutions during the period of writing (though no specific salary was set aside or given for the writing of this paper). No funding bodies had any role in the study design, data collection, analysis, decision to publish or preparation of the manuscript.

Competing Interests: Isabelle Boutron is a member of *PLOS Medicine* Editorial Board. The authors have declared that no other competing interests exist.

Abbreviation: RCT, randomized controlled trial

* E-mail: isabelle.boutron@hfd.aphp.fr

Introduction

The media play an important role in the dissemination of findings from health research. More than half of US adults report that they follow health news closely [1]. Further, 90% of the general public gets most of its information about science from the mass media [2]. Press releases are a major source of information for one-third of medical reports in US newspapers [3]. Press releases are widely used by the medical researchers to attract favorable media attention [4–6] and to promote their research [7–9]. A press release should provide journalists with the basic information needed to develop a news story and publish it in the mass media.

Randomized controlled trials (RCTs) are considered the gold standard for therapeutic evaluation [10]. Adequate and undistorted communication of the findings from RCTs is essential for physicians, researchers, and patients because it allows for efficient uptake of research into clinical practice [11]. Theoretically, in reports of RCTs published in peer-reviewed journals, the data should speak for themselves. However, a recent study showed that research findings can be distorted in published articles, by the use of “spin,” which is defined as specific reporting emphasizing the beneficial effect of the experimental treatment [12]. The types of distorted presentation or “spin” are diverse, with, for example, a particular focus on statistically significant results (within-group comparison, subgroup analyses, and secondary outcomes) or an inadequate interpretation of nonstatistically significant differences as demonstrating equivalence in treatment effectiveness or lack of difference in adverse events.

We aimed to (1) evaluate the presence of “spin” in press releases and associated media coverage and (2) evaluate whether findings of RCTs contained within press releases and media coverage are misinterpreted.

Methods

Selection of Press Releases, Related Scientific Articles, and News Items

We identified all press releases indexed in EurekAlert! (online free database for science press releases; www.eurekalert.org) between December 1, 2009, and March 31, 2010, using the following search strategy: topic “medicine and health,” type of release “research news,” keyword: random* [4,13]. We included press releases for published results of two-arm, parallel-group RCTs defined as prospective studies assessing health care interventions in human participants. To have a homogeneous sample, we excluded press releases for equivalence or noninferiority, cross-over, cluster, and multiple-arm trials; follow-up studies; press releases not reported in English; and those about more than one study. Duplicate press releases (i.e., press releases published more than once in the database) were systematically searched and excluded.

The title and full text of all retrieved press releases were screened by one reviewer to exclude any non-eligible press releases.

We obtained a copy of the scientific article related to the press release from (1) the direct link or full reference citation reported in the press release, if available; or (2) the PubMed single citation matcher indicating the year of publication, journal, and author’s name. Each retrieved scientific article (abstract and full text) was assessed by the same reader to confirm eligibility.

Finally, for all selected press releases, we systematically searched for related news items in the “general news” library of LEXIS-NEXIS using (1) the name of the disease; (2) the treatment being

evaluated, and, if needed, the name of the first or second author. All news related to the articles or press releases were retrieved, and we selected the news that had the highest number of words dedicated to the selected study.

Data Abstraction

Data were abstracted from the press release, news items, and the related published scientific article. For this purpose, we developed a standardized data-abstraction form using previous work on the same topics [12–14]. The data-abstraction form and details about the methods is available in Texts S1 and S2.

The data-abstraction form was preliminarily tested by two of the reviewers with a sample of 15 press releases and original articles indexed in January 2008. The data that involved some subjectivity, such as the type of “spin” were abstracted by two independent reviewers, with discrepancies resolved by consensus. Other data were evaluated by a single reviewer. The concordance between the two reviewers for the assessment of “spin” is reported in Text S3; the mean kappa coefficient for “spin” was 0.56 (range 0.43–0.69).

We systematically extracted data related to the characteristics of (1) the RCT, (2) the press releases, and (3) the presence of “spin” in the article abstract conclusions, in the press release and, when available, in the news items.

We defined “spin” as a specific reporting (intentional or unintentional) that emphasizes the beneficial effect of the experimental treatment. We used a classification of “spin” described in a previous work [12]. This classification was initially developed in the context of trials with a nonstatistically significant primary outcome. This classification was adapted for all RCTs. We considered “spin” as being a focus on statistically significant results (within-group comparison, secondary outcomes, subgroup analyses, modified population of analyses); an interpretation of statistically nonsignificant results for the primary outcomes as showing treatment equivalence or comparable effectiveness; or any inadequate claim of safety or emphasis of the beneficial effect of the treatment.

Results Interpretation

The RCT results were interpreted independently from three different sources: (1) from the full text of the scientific article, (2) from the press release, and (3) from the news items.

For each source, different pairs of assessors independently evaluated the results of the RCT and achieved consensus. Assessment based on the scientific article relied on the results for the primary outcomes, secondary outcomes, and harm. For assessment of press releases, assessors were blinded to the authors of the press release, the content of the scientific article, and the journal of publication. For assessment of news items, assessors were blinded to the content of the press release and scientific article. All results reported represent the consensus of each pair of assessors.

Interpreting the RCT results. The trial results were interpreted independently by use of the same scale, from 1 to 5 [15]. According to this scale, the assessors had to indicate whether patients should (1) definitely get the experimental treatment evaluated, (2) probably get the experimental treatment evaluated, (3) decide for themselves (i.e., the article was neutral), (4) probably not get the experimental treatment evaluated, or (5) definitely not get the experimental treatment evaluated. If the interpretation of the RCT results was classified as 1 or 2, the experimental treatment was considered beneficial; 3, the trial results were neutral; 4 or 5, the experimental treatment was considered not beneficial.

Definition of misinterpretation. Misinterpretation was defined as the interpretation of the press release or news items differing from that based on the full-text article by at least one class according to the above three-class system of scores. Misinterpretation of the press release or news items could overestimate the treatment beneficial effect or underestimate the treatment effect. For example, an overestimation of the treatment beneficial effect in the press release or news items occurred when reading the published article led to rating the trial results as neutral, whereas reading the press release or news items led to rating the experimental treatment as beneficial.

Statistical Analysis

Data for quantitative variables are expressed with medians and IQRs. Data for qualitative variables are expressed with frequencies and percentages. We planned bivariate and multivariable analysis to identify factors associated with (1) "spin" in the press releases, (2) an overestimation of the beneficial effect of the experimental treatment from press releases, (3) "spin" in the news items, and (4) an overestimation of the beneficial effect of the experimental treatment from news items. For bivariate analysis, we used the chi-square or Fisher exact test for categorical data and the Student *t*-test for quantitative data. For the multivariable analysis, we performed a Poisson regression with robust error variance [16] with a bootstrap model selection variable method [17] to assess all relevant variables. We used 1,000 bootstrap samples. Variables with $p < 0.25$ in bivariate analysis were selected for possible inclusion in the multivariable model. Variables identified as independent factors associated with "spin" in the press release in at least 60% of the bootstrap samples were kept in the multivariable model.

Results are expressed as risk ratio (RR) and 95% CIs. We did not perform multivariable analysis to identify factors associated with overestimation of the benefit of the experimental treatment because there were few events as compared with the number of variables to include.

Statistical analysis involved use of SAS v9.1 (SAS Institute).

Results

Selection of Press Releases and Scientific Articles

The search strategy in EurekaAlert! between December 1, 2009, and March 31, 2010, retrieved 498 press releases. The selection process resulted in 70 press releases and related scientific articles (Figure 1). Of these, 41 had associated news items. The list of press releases and published articles included is available in Text S4.

Characteristics of RCTS

The description of the scientific articles is in Table 1. In total, 38 (54%) articles were published in a specialized journal; the median (interquartile range) journal impact factor was 17.2 (4.8–28.4). The funding source was for-profit (only or with a nonprofit source) for about half of the reports. In 34 reports (49%), the primary outcomes were statistically significant, and in 24 (34%), all primary outcomes were not statistically significant. In all, 28 articles (40%) exhibited at least one type of "spin" in the abstract conclusions. The main types of "spin" in the abstract conclusions were no acknowledgement of nonstatistically significant primary outcomes (20%); interpreting $p > 0.05$ as demonstrating equivalence (7%); inappropriate extrapolation (9%); focus on statistically significant results such as subgroup analyses (6%), within-group comparisons (9%), and secondary outcomes (4%); or inadequate claim of safety (6%).

Characteristics of Press Releases

The general characteristics of press releases are in Table 2: 57% were written by a press officer; half provided easy access to the research article that had been press released (i.e., a direct link or the full reference) and 25 (36%) reported the funding source. The results for primary outcomes were reported with words only in 29 (41%) press releases. Safety was mentioned in 24 (34%) and quantified in 14 (20%); the study limitations were reported in ten (14%). A total of 58 (83%) press releases contained quotations from authors or editors of the article. In 30 (52%), the interview reported results with emphasis, such as "this work paves the way for further study," or in 22 (38%) with moderation, such as "further investigation is needed to establish (...)." Quotations from the article were included in 22 (31%) of the releases. In 11 (50%), the quotations reported results with emphasis, such as "clinical findings are indeed very encouraging, said Dr..."), or in seven (32%) with moderation.

About half of the press releases (33; 47%) had at least one type of "spin" (Table 2).

Factors Associated with "Spin" in Press Releases

From bivariate analysis (Table 3), "spin" in press releases was more frequent in trials published in a specialized journal (58% versus 34% in a general journal; $p = 0.05$), trials with small sample size (i.e., < 112) (63% versus 31%; $p = 0.008$), and trials with "spin" in the scientific article abstract conclusion (93% yes, versus 17% no; $p < 0.001$). The presence of "spin" in the press release was not associated with funding source (45% profit versus 49% other; $p = 0.8$), author of the press release (48% press officer versus 47% other; $p = 0.9$), the experimental treatment (47% drug versus 47% other; $p = 1.0$) or results of the primary outcome (46% all nonstatistically significant versus 48% other; $p = 0.9$). In multivariable analysis including all variables with $p < 0.25$ in the bivariate analysis (i.e., journal, "spin" in the abstract conclusion, and sample size), the only factor associated with "spin" in the press release was "spin" in the scientific article abstract conclusions (RR = 5.6, 95% CI 2.8–11.0, $p < 0.001$) (Text S5).

Interpretation of the Trial Results from Press Releases

For the interpretation based on the full-text scientific articles, for 38 articles (54%), the experimental treatment was considered beneficial, 18 (26%) neutral, and 14 (20%) not beneficial. In contrast, for the interpretation based on press releases, for 55 releases (79%), the experimental treatment was considered beneficial, two (3%) neutral, and 13 (18%) not beneficial. The results were misinterpreted in 22 press releases (31%); for 19 (86%), the assessors overestimated the benefit of the experimental treatment from the press release and for three (14%), they underestimated the benefit of the experimental treatment from the press release.

As shown in Table 4, on the basis of press releases, the benefit of the experimental treatment was overestimated more often for trial results published in a specialized journal rather than in a general medical journal (45% versus 6%; $p < 0.001$), for trials with a small rather than large sample size (46% versus 9%; $p < 0.001$), for trials with nonstatistically rather than significant primary outcomes (42% versus 20%; $p = 0.05$), and for trials with "spin" rather than without "spin" in the press release (48% versus 8%; $p < 0.001$). These results did not differ significantly by funding source, author of the press release, or type of experimental treatment.

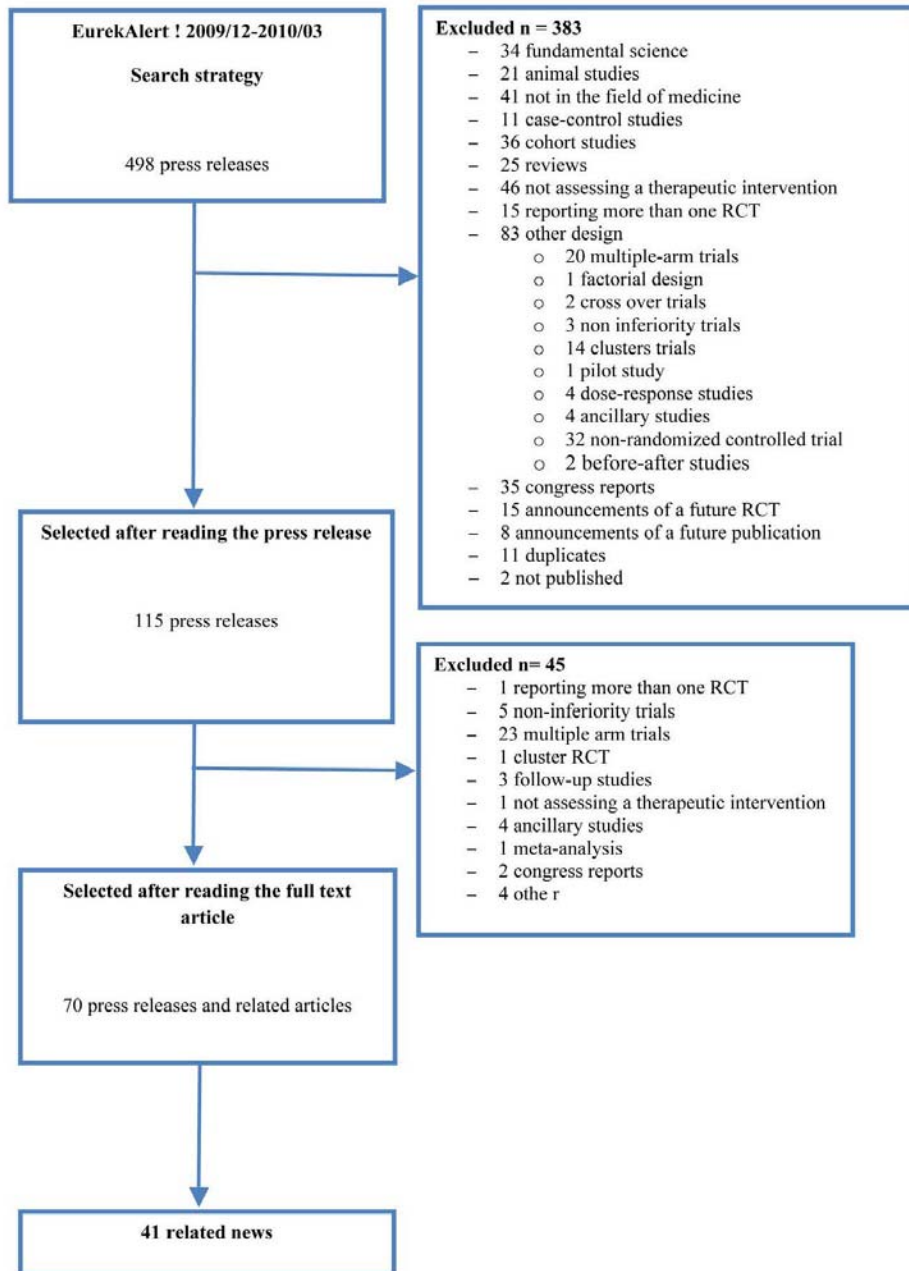


Figure 1. Flow diagram of the selected press releases and related articles.
doi:10.1371/journal.pmed.1001308.g001

“Spin” and Interpretation of the News

For a sample of 41 RCTS we retrieved the scientific article, the press release, and any news items. “Spin” was identified in 17 (41%) abstracts, 19 (46%) press releases, and 21 (51%) news items.

Figure 2 describes the reporting of “spin” in abstracts, press releases, and news items. For the 17 abstracts reported with “spin”, 16 press releases and related news items featured the same “spin.” For the 24 abstracts without “spin,” only three press releases featured “spin,” which was subsequently reported in the

related news items. Examples of “spin” in the abstract and related press releases and news items are in Figure 3. The factors associated with “spin” in the news were specialty journals (67% versus 35%; $p=0.04$), small sample size (68% versus 32%; $p=0.02$), “spin” in abstract (100% versus 5%; $p<0.001$), and “spin” in the press release (100% versus 13%; $p<0.001$) (Text S6).

Overall, the assessors overestimated the benefit of the experimental treatment from the news for 10 (24%) reports. Factors associated with overestimation of the beneficial effect of the treatment from the news items were small sample size (41% versus

Table 1. General characteristics of articles.

Characteristics	Subcharacteristics	n=70 (%)
Type of journal, n (%)	General medical journal	32 (46)
	Specialized medical journal	38 (54)
Funding source, n (%)	Profit or both profit and nonprofit	33 (47)
	None or nonprofit	33 (47)
	Not reported	4 (6)
Sample size median; [IQR]; (min-max)		112; [54–435]; (16–94,370)
Experimental treatment, n (%)	Drug	36 (51)
	Surgery/procedure	9 (13)
	Device	5 (7)
	Therapeutic strategy	7 (10)
	Participative intervention	12 (17)
Comparator, n (%)	Other	1 (1)
	Placebo	29 (41)
	Active treatment	32(46)
Primary outcomes clearly identified, n (%)	Other	9 (13)
		61(87)
		61 (87)
Type of primary outcomes, n (%)	Efficacy	61 (87)
	Safety	1 (1)
	Both	4 (6)
	Unclear	4 (6)
Primary outcomes reported adequately, n (%) ^a		56 (80)
Results of primary outcomes, n (%)	All statistically significant	34 (49)
	All statistically nonsignificant	24 (34)
	Some statistically significant/some not	11 (16)
	Unclear	1 (1)
At least one “spin”		28 (40)
Type of “spin” ^b	No acknowledgment of nonstatistically significant primary outcome	14 (20)
	Claiming equivalence when results failed to demonstrate a statistically significant difference	5 (7)
	Focus on positive secondary outcome	3 (4)
	Focus on inappropriate subgroup	4 (6)
	Focus on within-group (or over-all within) comparison	6 (9)
	Nonstatistically significant outcome reported as if they were significant	3 (4)
	Ignored data of safety	1 (1)
	Inadequate claim of safety	4 (6)
	Inappropriate extrapolation	6 (9)
	Other	5 (7)

^aAdequately, with effect size and precision or treatment effect in each arm with precision.

^bNumbers do not add up as the types of “spin” were not mutually exclusive.

doi:10.1371/journal.pmed.1001308.t001

Table 2. General characteristics of press releases.

Characteristics	Subcharacteristics	n=70 (%)
Origin, n (%)	Press officer	40 (57)
	Industry or institution	30 (43)
Easy access to full article, (i.e., direct link or the full reference) n (%)		36 (51)
Funding reported, n (%)		25 (36)
Design reported, n (%)		70 (100)
Sample size reported, n (%)		65 (93)
Length of follow-up reported, n (%)		46 (66)
Primary outcomes reported, n (%)	In words only	29 (41)
	Per arms	30 (43)
	With effect size	17 (24)
Safety reported, n (%)	Mentioned	24 (34)
	Quantified	14 (20)
Limits reported, n (%)		10 (14)
Interview included, n (%)	Authors only	40 (57)
	Experts or editorialists only	6 (9)
	Both	12 (17)
Article quotation reported, n (%)		22 (31)
At least one type of "spin"		33 (47)
Type of "spin" ^a	No acknowledgment of nonstatistically significant primary outcome	13 (19)
	Claiming equivalence when results failed to demonstrate a statistically significant difference	7 (10)
	Focus on positive secondary outcome	5 (7)
	Focus on inappropriate subgroup	4 (5)
	Focus on within-group (or over-all within) comparison	11 (16)
	Nonstatistically significant outcome reported as if they were significant	5 (7)
	Ignored data of safety	3 (4)
	Inadequate claim of safety	5 (7)
	Inappropriate extrapolation	6 (9)
	Other "spin"	2 (3)

^aNumbers do not add up as the types of "spin" were not mutually exclusive.
doi:10.1371/journal.pmed.1001308.t002

Table 3. Bivariate analysis of factors associated with and "spin" in the press releases.

Characteristics	Subcharacteristics	"Spin" in Press Release n/Total n (%)	p-Value
Journal	General	11/32 (34)	0.05
	Specialized	22/38 (58)	—
Funding source	Profit	15/33 (45)	0.8
	Nonprofit or not reported	18/37 (49)	—
Sample size	<112	22/35 (63)	0.008
	≥112	11/35 (31)	—
Experimental treatment	Drug	17/36 (47)	1.0
	Other	16/34 (47)	—
Results of primary outcome(s)	All nonstatistically significant	11/24 (46)	0.9
	Other	22/46 (48)	—
Authors of press release	Press officer	19/40 (48)	0.9
	Other	14/30 (47)	—
"Spin" in abstract conclusion	Yes	26/28 (93)	<0.001
	No	7/42 (17)	—

doi:10.1371/journal.pmed.1001308.t003

Table 4. Bivariate analysis of factors associated with an overestimation of the benefit of the experimental treatment from the press releases as compared with the interpretation from articles.

Characteristics	Subcharacteristics	Overestimation of the Benefit of the Experimental Treatment n/Total n (%)	p-Value
Journal	General	2/32 (6)	<0.001
	Specialized	17/38 (45)	
Funding source	Profit	7/33 (21)	0.3
	Nonprofit or not reported	12/37 (32)	
Sample size	n<112	16/35 (46)	<0.001
	n≥112	3/35 (9)	
Experimental treatment	Drug	11/36 (31)	0.5
	Other	8/34 (24)	
Results of primary outcome(s)	All nonstatistically significant	10/24 (42)	0.05
	Other	9/46 (20)	
Authors of press release	Press officer	10/40 (25)	0.6
	Other	9/30 (30)	
"Spin" in press releases	Yes	16/33 (48)	<0.001
	No	3/37 (8)	

doi:10.1371/journal.pmed.1001308.t004

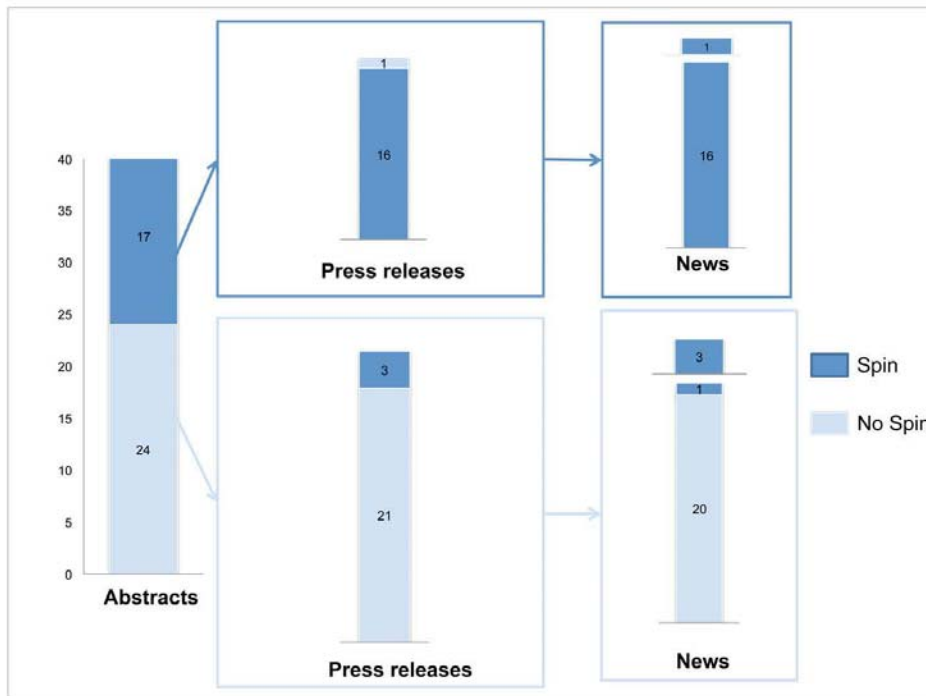


Figure 2. "Spin" in abstract conclusions, press releases, and news items.
doi:10.1371/journal.pmed.1001308.g002

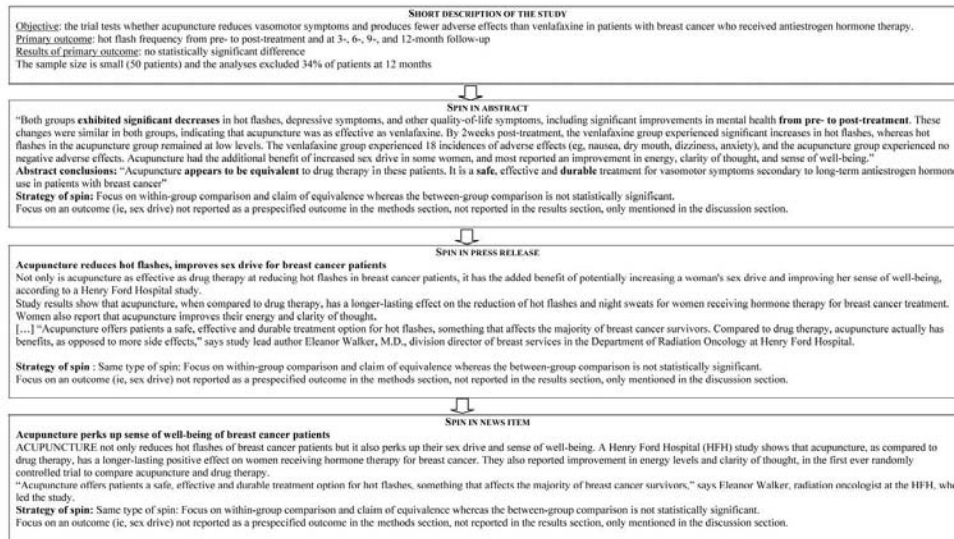


Figure 3. Examples of "spin" in abstracts, in press releases, and in related news items.
doi:10.1371/journal.pmed.1001308.g003

5%, $p = 0.01$), and "spin" in the news (43% versus 5%, $p = 0.009$) (Text S7).

Discussion

Our results highlight a tendency for press releases and the associated media coverage of RCTs to place emphasis on the beneficial effects of experimental treatments. This tendency is probably related to the presence of "spin" in conclusions of the scientific article's abstract. This tendency, in conjunction with other well-known biases such as publication bias, selective reporting of outcomes, and lack of external validity, may be responsible for an important gap between the public perception of the beneficial effect and the real effect of the treatment studied.

Previous studies have highlighted the importance of press releases for results communication and dissemination [4,5]. Indeed, as a direct means of communication between medical journals and the media, press releases provide an opportunity for journals to influence how the research is translated into news [4]. The press release is essential when considering the impact of press coverage by the media on health care utilization, clinical practice, and researchers' behavior [1]. This influence has been clearly shown in a quasi-experimental study evaluating the impact of media coverage [11]. The authors compared the number of scientific citations of articles published in the *New England Journal of Medicine* that were covered by the *New York Times* to similar articles that were not covered. They also performed this comparison during a 3-mo period when the *New York Times* was on strike; the *New York Times* continued to print an "edition of record" but did not sell copies to the public because of the strike. The authors demonstrated that the high citation of articles covered by the *New York Times* was not present during the strike. Consequently, the high citation was related to the media coverage, not the

importance of the research [18]. A Cochrane systematic review highlighted the impact of the mass media on health services utilization [19]. It showed a consistent effect after planned campaigns and unplanned coverage. Another study showed a clear association of the media coverage of invasive group A streptococcal (GAS) disease and testing for GAS in pediatric emergency departments, with an important increase in the prescription of rapid tests for GAS in pediatric emergency departments concomitant with a peak in media attention, despite no increase in the number of children presenting symptoms that might warrant such testing [20].

Unfortunately, as shown in our study, and previous work the quality of media reports is questionable. An assessment of the reporting of medical news in the mainstream media highlighted the inadequate accuracy and balance of the news media in reporting medical science [21–23]. The criticisms of the mainstream media also applied to press releases. Woloshin et al., in evaluating press releases issued by 20 academic medical centers, showed that the releases frequently promoted preliminary research without giving basic details or the cautions needed to judge the meaning, relevance, or validation of the science (42% of press releases evaluated in this study did not provide any relevant caveats, and 90% about animal or laboratory studies lacked caveats about extrapolating results to humans) [13]. Furthermore, press releases tended to overstate the importance of the research, 29% were rated as exaggerating the findings' importance and 26% of investigator quotes were considered to overstate the research importance [13]. Recently, a study showed that the quality of press releases influenced subsequent media coverage content [24].

Of course, press releases are not meant to be condensed versions of scientific papers; they are meant to summarize the most important findings, contextualize these findings for journalists, and provide contact details for authors and quotes. By being

condensed, they always lack details that are contained in the papers. The use of “spin” or a particular emphasis could be a way to increase the interest of journalists and subsequent citations in the peer-reviewed literature.

However, this situation becomes problematic if it modifies readers’ interpretation of research findings. Our results add to these previous studies by showing the link between the distorted presentation and interpretation of the results in scientific articles and the distorted content and interpretation of press releases. These findings raise the issue of the quality of the peer review process and highlight the importance of this process for disseminating accurate research results.

Our study has several limitations. Firstly, our sample included only published reports of RCTs with a press release indexed in the EurekaAlert! database within a 4 mo period, and reported in English; this sample may not be representative of all press releases of RCT results. In fact, half of the press releases selected were written by press officers of medical journals with a high impact factor. Other sources of press releases exist on industry websites, medical journal websites, or other databases for journalists. However, the EurekaAlert! database is one of the most important sources of freely available press releases, and most research published on press releases has used this database. Further, there is no reason to believe that the selection of the sample over only 4 mo would bias the results. Secondly, RCTs represent only a small part of the medical literature and the findings may not apply to media reporting of medical or scientific research as a whole. Thirdly, we searched for “spin” only in the article abstract conclusions, not in the entire published article. Consequently, we are not able to determine whether “spin” in the press release was the same as the “spin” in the whole article. We chose the abstract conclusions because it is the most accessible section of an article. Readers often base their initial assessment of a trial on the information reported in an abstract conclusion, and in some geographic areas, the abstract of an RCT report may be all that health professionals have easy access to [25,26]. Fourthly, the content analysis and the interpretation coding were subjective [27]. However, two independent reviewers performed this assessment with consensus. Fifthly, we focused on articles and press releases of RCT results. We did not evaluate press releases for other study designs or proceedings of conferences.

In conclusion, previous work showed that exaggerated and inappropriate coverage of research findings in the news media is linked to inappropriate reporting of press releases. Our study adds to these results showing that “spin” in press releases and the news is related to the presence of “spin” in the published article, namely the abstract conclusions. Additionally, our work highlights that this

inappropriate reporting could bias readers’ interpretation of research results.

Consequently, reviewers and editors of published articles have an important role to play in the dissemination of research findings and should be particularly aware of the need to ensure that the conclusions reported are an appropriate reflection of the trial findings and do not overinterpret or misinterpret the results.

Supporting Information

Text S1 Data abstraction form.
(DOC)

Text S2 Details related to the method.
(DOC)

Text S3 Kappa coefficient or agreement percentage for the assessment of “spin” in press releases and in articles.
(DOC)

Text S4 List of press releases and published articles examined.
(DOC)

Text S5 Multivariate analysis of factors associated with “spin” in press releases.
(DOC)

Text S6 Bivariate analysis of factors associated with “spin” in news items ($n = 41$).
(DOC)

Text S7 Bivariate analysis of factors associated with an overestimation of the benefit of the experimental treatment from the news as compared with the interpretation from the article abstract conclusions ($n = 41$).
(DOC)

Acknowledgments

We would like to thank Elodie Perrodeau, Nizar Ahmad, Karima Amazzough, and Guillaume Lonjon for their help in this study.

Author Contributions

Conceived and designed the experiments: AY IB PR. Performed the experiments: AY AB IM PC. Analyzed the data: AY. Contributed reagents/materials/analysis tools: AY IB AB IM PC JM PR. Wrote the first draft of the manuscript: AY IB. Contributed to the writing of the manuscript: AY IB PR. ICMJE criteria for authorship read and met: AY IB AB IM PC JM PR. Agree with manuscript results and conclusions: AY IB AB IM PC JM PR.

References

- Schwartz LM, Woloshin S (2004) The media matter: a call for straightforward medical reporting. *Ann Intern Med* 140: 226–228.
- Jaques H (2011) *BMJ Careers* - get your research reported well in the news. Available: <http://careers.bmj.com/careers/advice/view-article.html?id=20001805>. Accessed 8 August 2012.
- (2009) Incomplete reporting of research in academic press releases. *Lancet* 373: 1920.
- Woloshin S, Schwartz LM (2002) Press releases: translating research into news. *JAMA* 287: 2856–2858.
- van Trigt AM, de Jong-van den Berg LT, Haaijer-Ruskamp FM, Willems J, Tromp TF (1994) Journalists and their sources of ideas and information on medicines. *Soc Sci Med* 38: 637–643.
- Kuriya B, Schneid EC, Bell CM (2006) Quality of pharmaceutical industry press releases based on original research. *PLoS One* 3: e2828. doi:10.1371/journal.pone.0002828
- Chapman S, Nguyen TN, White C (2007) Press-released papers are more downloaded and cited. *Tob Control* 16: 71.
- Entwistle V (1995) Reporting research in medical journals and newspapers. *BMJ* 310: 920–923.
- Woloshin S, Schwartz LM, Kramer BS (2009) Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 101: 1596–1599.
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312: 71–72.
- Phillips DP, Kanter RJ, Bednarczyk B, Tastad PL (1991) Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 325: 1180–1183.
- Boutron I, Dutton S, Ravaud P, Altman DG (2010) Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 303: 2058–2064.
- Woloshin S, Schwartz LM, Casella SL, Kennedy AT, Larson RJ (2009) Press releases by academic medical centers: not so academic? *Ann Intern Med* 150: 613–618.
- Als-Nielsen B, Chen W, Gluud C, Kjaergard LL (2003) Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 290: 921–928.
- Schwartz LM, Woloshin S (2002) News media coverage of screening mammography for women in their 40 s and tamoxifen for primary prevention of breast cancer. *JAMA* 287: 3136–3142.

16. Zou G (2004) A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 159: 702–706.
17. Austin PC, Tu JV (2004) Bootstrap methods for developing predictive models. *The American Statistician* 58: 131–137.
18. Haas JS, Kaplan GP, Gerstenberger EP, Kerlikowske K (2004) Changes in the use of postmenopausal hormone therapy after the publication of clinical trial results. *Ann Intern Med* 140: 184–188.
19. Grilli R, Ramsay C, Minozzi S (2002) Mass media interventions: effects on health services utilisation. *Cochrane Database Syst Rev*: CD000389.
20. Sharma V, Dowd MD, Swanson DS, Slaughter AJ, Simon SD (2003) Influence of the news media on diagnostic testing in the emergency department. *Arch Pediatr Adolesc Med* 157: 257–260.
21. Schwartz LM, Woloshin S, Welch HG (1999) Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med* 341: 277–279.
22. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, et al. (2000) Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 342: 1645–1650.
23. Naylor CD, Chen E, Strauss B (1992) Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 117: 916–921.
24. Schwartz LM, Woloshin S, Andrews A, Stukel TA (2012) Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study. *BMJ* 344: d8164.
25. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, et al. (2008) CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5: e20. doi:10.1371/journal.pmed.0050020
26. Gøtzsche PC (2006) Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ* 333: 231–234.
27. Horton R (2002) The hidden research paper. *JAMA* 287: 2775–2778.

3. Impact de l'ajout d'une section limite dans les résumés des revues systématiques sur l'interprétation des lecteurs

3.1. Justification de l'étude

Les revues systématiques sont la pierre angulaire de l'évaluation thérapeutique. Cependant leur interprétation est difficile[51,52]. Les recommandations « PRISMA for abstract » recommandent que les limites de la revue systématique soient rapportées dans le résumé de celle-ci[25]. Aucune étude n'a à ce jour étudié l'impact de cette section limite écrite selon les recommandations PRISMA sur l'interprétation des résultats par les lecteurs.

3.2. Objectif

L'objectif de ce deuxième travail de thèse était d'évaluer l'impact de l'ajout d'une section limite dans les résumés des revues systématiques sur l'interprétation des résultats par les lecteurs.

3.3. Méthode

Nous avons réalisé un essai randomisé contrôlé pour répondre à cette question.

Dans un premier temps nous avons sélectionné 30 résumés de revues systématiques à partir d'un échantillon représentatif de 100 revues non Cochrane indexées dans DARE (DAAtabase of Reviews of Effets) entre janvier et mars 2012[134]. Nous avons inclus les revues systématiques étudiant une intervention de santé, qui rapportaient une évaluation des risques de biais et dont la conclusion du résumé était en faveur de l'effet bénéfique de l'intervention

expérimentale.

Après lecture du texte intégral de la revue, nous avons écrit pour chacune des revues sélectionnées, une section limite rapportant les limites de l'étude selon les recommandations « PRISMA for abstract »[25]. Ces limites concernaient les 5 points spécifiés dans les recommandations PRISMA : 1) « the risk of bias », le risque de biais des études incluses dans la revue systématique, 2) « the inconsistency of effect or association », c'est-à-dire l'hétérogénéité importante entre les études incluses, 3) « the imprecision », l'imprécision liée au faible nombre d'évènements ou aux petits effectifs, 4) « the indirectness », c'est à dire l'utilisation d'un critère intermédiaire ou à court terme, 5) « the likely publication bias », c'est-à-dire le biais de publication potentiel.

Nous avons ainsi constitué deux groupes de résumés. Le groupe expérimental contenait 30 résumés modifiés par l'ajout d'une section rapportant les limites de la revue. Le groupe contrôle contenait 30 résumés sous leur format original. Les noms des auteurs, le nom des interventions et le journal de publication de la revue systématique étaient masqués.

Les participants potentiels étaient des auteurs correspondants d'essais cliniques publiés entre janvier 2010 et juin 2013 et indexé dans MEDLINE. Les participants potentiels étaient invités par email à participer à une étude en ligne sur l'interprétation des résumés des revues systématiques. En cas d'acceptation, le participant était alors dirigé vers un site internet sécurisé, où seuls les cliniciens étaient inclus (les copies d'écran du site sont visibles dans l'annexe 1). Après inclusion, les participants étaient randomisés dans un des deux groupes (avec ou sans section limite) et étaient invités à lire un résumé de revue systématique puis à répondre à 5 questions. Le critère de jugement principal était la confiance du lecteur dans les résultats de la revue systématique présentée dans le résumé. Les critères secondaires étaient la

confiance dans 1) la validité de la conclusion, 2) l'effet bénéfique de l'intervention expérimentale, 3) la capacité des résultats à modifier la pratique clinique, 4) la rigueur de la revue systématique. Chaque question était cotée de 0 « pas confiance du tout » à 10 « très confiant ». Les participants étaient aveugles des objectifs de l'étude et de l'existence des deux groupes de résumés avec ou sans limites.

L'unité d'analyse était le résumé, 266 participants étaient nécessaires pour détecter une différence de 0,4 point avec une puissance de 90% et un risque alpha de 5 %. Nous avons choisi d'inclure 300 participants afin que chaque résumé avec ou sans limite soit lu par 5 participants.

3.4. Résultats

Parmi les 4807 participants potentiels invités à participer à notre étude, 394 se sont connectés au site, 89 ont été exclus parce qu'ils n'étaient pas cliniciens et 5 n'ont pas complété le questionnaire. 300 participants ont été inclus, 150 randomisés dans le groupe expérimental « résumés avec section limites » et 150 randomisés dans le groupe contrôle « résumés sans section limite ».

Les participants étaient des « lecteurs – experts », la moitié (53%) avait été impliquée au moins une fois dans la réalisation d'une revue systématique et 59 ont déclaré lire régulièrement des revues systématiques. La moitié des participants était localisée en Europe et un tiers aux Etats-Unis.

Parmi les 30 résumés sélectionnés, 23 étaient des résumés structurés et sept résumés étaient

non structurés. La section limite était ajoutée avant la conclusion, avec un en-tête intitulé « limites » pour les résumés structurés, et directement intégré dans le texte juste avant la phrase de conclusion et sans en-tête pour les résumés non structurés (voir exemple tableau 6). De plus, 9 résumés rapportaient initialement au moins une limite que nous avons supprimée et remplacée par la section limite que nous avons écrite selon les recommandations PRISMA. Les 30 sections limites créées concernaient le risque de biais des études incluses dans 73% (n=22), l'hétérogénéité dans 43% (n=13), le biais de publication dans 50% (n=15), l'imprécision dans 40% (n=12) et l'« indirectness » dans 7% (n=2) des résumés.

L'évaluation par les lecteurs des résumés avec ou sans section limite n'était pas différente pour le critère de jugement principal et pour les critères de jugements secondaires. Les résultats sont présentés dans le tableau 6 ci-dessous.

Tableau 5: Exemple de résumé avec et sans section limite

Exemple de résumé sans section limite	Exemple de résumé avec section limite
<p>Title:</p> <p>Quality of life and related dimensions in cancer patients treated with intervention A: a meta-analysis.</p> <p>Objectives:</p> <p>The aim of this meta-analysis was to determine the effectiveness of the intervention A, in the treatment of patients with cancer with respect to quality-of-life- (QoL-) associated measures.</p> <p>Methods:</p> <p>We searched databases such as PubMed/Medline, ExcerptaMedica Database (EMBASE), CAMbase, and other for controlled clinical studies on parameters associated with QoL. Outcome data were extracted and converted into standardized mean differences and their standard errors.</p> <p>Results:</p> <p>Thirteen prospective and controlled studies which met the inclusion/exclusion criteria reported positive effects in favor of the intervention A. A random-effect meta-analysis estimated the overall treatment effect at standardized mean difference = 0.56 (CI: 0.41 to 0.71, P < .0001).</p> <p>Conclusions:</p> <p>The analyzed studies give some evidence that intervention A might have beneficial short-time effects on QoL-associated dimensions and psychosomatic self-regulation</p>	<p>Title:</p> <p>Quality of life and related dimensions in cancer patients treated with intervention A: a meta-analysis.</p> <p>Objectives:</p> <p>The aim of this meta-analysis was to determine the effectiveness of the intervention A, in the treatment of patients with cancer with respect to quality-of-life- (QoL-) associated measures.</p> <p>Methods:</p> <p>We searched databases such as PubMed/Medline, ExcerptaMedica Database (EMBASE), CAMbase, and other for controlled clinical studies on parameters associated with QoL. Outcome data were extracted and converted into standardized mean differences and their standard errors.</p> <p>Results:</p> <p>Thirteen prospective and controlled studies which met the inclusion/exclusion criteria reported positive effects in favor of the intervention A. A random-effect meta-analysis estimated the overall treatment effect at standardized mean difference = 0.56 (CI: 0.41 to 0.71, P < .0001).</p> <p>Limitations:</p> <p>This systematic review is limited by the very poor methodological quality of the included studies. Further, there is some evidence of publication bias.</p> <p>Conclusions:</p> <p>The analyzed studies give some evidence that intervention A might have beneficial short-time effects on QoL-associated dimensions and psychosomatic self-regulation.</p>

Tableau 6: Résultats du critère principal et des critères secondaires

	Résumés sans section limites Moyenne (ET) n=150	Résumés avec section limites Moyenne (ET) n=150	Différence moyenne (95% IC)	p
Quelle est votre confiance dans les résultats de cette étude (0-10)?	4,6 (2,5)	4,4 (2,3)	0,2 (-0,4 à 0,7)	p = 0,5
Quelle confiance avez-vous dans la validité de la conclusion de cette étude (0-10)?	4,1 (2,5)	4,0 (2,3)	0,07 (-0,5 à 0,6)	p = 0,8
Avez vous confiance dans le fait que l'intervention A soit bénéfique pour les patients (0-10)?	4,4 (2,6)	4,3 (2,3)	0,1 (-0,4 à 0,7)	p = 0,6
Avez vous confiance dans le fait que les résultats de cette étude puissent modifier votre pratique clinique (0-10) ?	3,8 (2,6)	3,8 (2,3)	-0,08 (-0,6 à 0,5)	p = 0,8
Pensez vous que cette étude a été réalisée rigoureusement (0-10) ?	4,1 (2,7)	4,4 (2,6)	-0,4 (-1,0 à 0,2)	p = 0,2

3.5. Discussion

Nous avons réalisé un essai randomisé contrôlé impliquant un grand nombre de lecteurs experts dans le domaine de la recherche clinique. Ce type de plan expérimental est très rarement utilisé en recherche méthodologique ou pour l'étude des facteurs influençant l'interprétation des résultats par les lecteurs. Pourtant il s'agit du plan expérimental apportant le meilleur niveau de preuve car il permet de contrôler au mieux les biais potentiels. De plus notre étude est originale car contrairement à d'autres études expérimentales dans ce domaine nous avons utilisé des résumés de revues systématiques publiées et indexées et non pas des

résumés fictifs[63].

Dans notre étude, la présence d'une section limite dans le résumé des revues systématiques ne semble pas avoir un impact sur l'interprétation des résultats par des lecteurs experts.

Malgré la sélection d'un échantillon de revues systématiques de bonne qualité, le niveau de confiance accordé par les lecteurs était relativement faible. Aussi l'ajout d'informations supplémentaires portant sur les limites de la revue n'a pas eu d'incidence sur l'interprétation globale des résultats par des lecteurs-experts.

On peut supposer que les lecteurs-experts interprètent une méta-analyse avec les limites et les biais potentiels de ce type d'étude. Le haut niveau d'expertise des participants à notre étude pourrait expliquer au moins en partie ces résultats. En effet plus de la moitié des participants avait déjà eu une expérience soit dans la réalisation d'une revue systématique soit dans la relecture de revues systématiques pour des journaux à comité de lecture. Ce niveau d'expertise pourrait augmenter leur prise de conscience des limites communément retrouvées dans les revues systématiques telles que le risque de biais de publication ou le risque de biais des études incluses dans la revue. En outre, la section limite créée était rapportée de façon factuelle et neutre, et nous n'avons pas modifié la conclusion du résumé. L'évaluation de la confiance dans les résultats d'une revue systématique est complexe car elle dépend à la fois de la réalisation de la revue systématique et de la qualité des études incluses dans la revue. Il y a donc un double de niveau de jugement de la qualité.

Notre étude comporte plusieurs limites. Premièrement les participants n'avaient pas accès au texte intégral de la revue mais uniquement au résumé. Deuxièmement, les participants avaient un bon niveau d'expertise et ne sont pas représentatifs de tous les lecteurs de revues systématiques. Par conséquent nous ne pouvons pas exclure que l'ajout de la section limite

pourrait être utile pour des lecteurs moins aguerris. Et dernièrement, nous avons exploré uniquement l'impact de la section limite dans les résumés des publications rapportant une revue systématique. De ce fait, nous ne pouvons pas extrapoler ces résultats aux publications rapportant d'autres types d'études.

Notre étude ne permet pas de faire des nouvelles recommandations pour la rédaction des revues systématiques ou de modifier celles en vigueur actuellement. En effet, cette étude est la première à avoir étudié l'impact de la section limite sur l'interprétation des lecteurs, ces résultats doivent être confirmés. De plus d'autres études impliquant des participants moins experts ou d'autre type de plans expérimentaux sont nécessaires.

3.6. Article

Yavchitz A, Ravaud P, Hopewell S, Baron G, Boutron I. Impact of adding a limitations section to abstracts of systematic reviews on readers' interpretation: a randomized controlled trial. *BMC Med Res Methodol.* 2014 Nov 24;14:123. doi: 10.1186/1471-2288-14-123.

Ce travail a été présenté en session orale au 7^{ème} congrès international du « Peer Review and Biomedical Publication Congress » en septembre 2013, Chicago, USA.

RESEARCH ARTICLE

Open Access

Impact of adding a limitations section to abstracts of systematic reviews on readers' interpretation: a randomized controlled trial

Amélie Yavchitz^{1,2,3,4*}, Philippe Ravaud^{1,2,3,5}, Sally Hopewell^{1,3,6}, Gabriel Baron^{1,2,3} and Isabelle Boutron^{1,2,3}

Abstract

Background: To allow an accurate evaluation of abstracts of systematic reviews, the PRISMA Statement recommends that the limitations of the evidence (e.g., risk of bias, publication bias, inconsistency, imprecision) should be described in the abstract. We aimed to evaluate the impact of adding such limitations sections on reader's interpretation.

Method: We performed a two-arm parallel group randomized controlled trial (RCT) using a sample of 30 abstracts of systematic reviews evaluating the effects of healthcare intervention with conclusions favoring the beneficial effect of the experimental treatments. Two formats of these abstracts were derived: one reported without and one with a standardized limitations section written according to the PRISMA statement for abstracts. The primary outcome was readers' confidence in the results of the systematic review as stated in the abstract assessed by a Likert scale from 0, not at all confident, to 10, very confident. In total, 300 participants (corresponding authors of RCT reports indexed in PubMed) were randomized by a web-based randomization procedure to interpret one abstract with a limitations section (n = 150) or without a limitations section (n = 150). Participants were blinded to the study hypothesis.

Results: Adding a limitations section did not modify readers' interpretation of findings in terms of confidence in the results (mean difference [95% confidence interval] 0.19 [-0.37-0.74], p = 0.50), confidence in the validity of the conclusions (0.07 [-0.49-0.62], p = 0.80), or benefit of the experimental intervention (0.12 [-0.42-0.44], p = 0.65). This study is limited because the participants were expert-readers and are not representative of all systematic review readers.

Conclusion: Adding a limitations section to abstracts of systematic reviews did not affect readers' interpretation of the abstract results. Other studies are needed to confirm the results and explore the impact of a limitations section on a less expert panel of participants.

Trial registration: ClinicalTrials.gov (NCT01848782).

Keywords: Meta-analysis, Systematic review, Bias, Limits, Limitation, Interpretation, Interpretation bias, Misinterpretation, Abstract, Results

* Correspondence: ayavchitz@gmail.com

¹Centre de Recherche Épidémiologies et Statistiques INSERM U1153, Paris, France

²Centre of Clinical Epidemiology, Assistance Publique des Hôpitaux de Paris, Hôtel Dieu Hospital, Place du Parvis Notre-Dame, 75181 Paris, Cedex 4, France
Full list of author information is available at the end of the article



© 2014 Yavchitz et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. 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

Systematic reviews are the cornerstone of therapeutic evaluation [1]. Clinicians, decision makers and researchers use them to keep up-to-date with current medical literature, develop clinical practice guidelines and sometimes plan future research [2-4]. However, systematic reviews may differ in quality depending on the conduct of the systematic review as well as the availability and quality of the primary studies [5]. Consequently, readers should carefully examine the methodological quality of reviews to assess their confidence in the results and conclusions.

Clinical decision makers and healthcare practitioners frequently rely on abstracts to decide the value of a study [6,7]. In some cases, healthcare practitioners have access to only the abstract and make healthcare decisions based solely on the information in the abstract [8]. To improve the transparency of abstracts [9,10], methodologists, researchers and editors have established recommendations for the presentation of systematic reviews in the PRISMA statements [11] with an extension for reporting abstracts [12]. These recommendations indicate that the limitations of the evidence should be described in the abstracts of systematic reviews [12]. This recommendation is rarely implemented [13], although some journals request that authors report a structured abstract with a limitations section [14]. However, the impact of adding this section to abstracts on readers' interpretation is unknown.

We aimed to evaluate the impact of adding a limitations section, written according to the PRISMA statement for abstracts of systematic reviews, on readers' interpretation of the abstract results.

Methods

Study design

We planned a randomized controlled trial to compare the interpretation of systematic review abstracts reported with and without a limitations section. We selected 30 abstracts of systematic reviews, and then developed 2 formats of the selected abstracts: reported with and without a limitations section written according to the PRISMA statement for abstracts. Then, we randomized participants to read and interpret one abstract with or without a limitations section. The study reporting follows the 2010 CONSORT statement [15].

Selection of abstracts of systematic reviews

The abstracts were selected from a cohort of 100 systematic reviews assessing the effects of healthcare interventions, published between January and March 2012, and indexed in the Database of Reviews of Effects [13]. The search strategy and eligibility criteria for this cohort were described elsewhere [13]. From this cohort, we selected the first 30 systematic reviews whose abstract conclusion favored the beneficial effect of the experimental

intervention and reported the risk of bias of primary studies in the full-text article, thus allowing selection of a relatively homogeneous sample. Abstracts from the cohort of 100 systematic reviews were screened by one author (AY) and reviewed by a second author (IB) to confirm eligibility. Any disagreements were resolved by consensus.

Construction of the limitations section

According to the PRISMA statement for abstracts, the limitations section should address the following: 1) risk of bias common to many or all studies, such as lack of blinding for subjective outcomes or unavailability of data; 2) inconsistency of effect or association, as demonstrated by high heterogeneity; 3) imprecision due to few events or small sample size, for example; 4) indirectness of the evidence, such as use of an intermediate or short-term outcome; and 5) likelihood of publication bias [12,16]. These limitations are the factors used to evaluate the level of evidence according to the GRADE approach [17].

One author (AY) systematically searched for and extracted from the full-text systematic review the following: 1) whether the systematic review had one or several of the limitations described above and 2) the limitations were reported by the authors of the systematic review in the full-text article or in the abstract. When limitations were outcome-specific, the limitations reported referred to the outcomes highlighted in the abstracts.

Then, for each selected abstract, one of the authors (AY) wrote a limitations section. The section focused on the limitations identified and was written in a standardized way, beginning with "This review is limited by....", with a maximum of 2 sentences. When the original abstract was structured, we preserved the structured form and added the limitations section before the conclusions section with a heading "Limitations". When the original abstract was not structured, we added the limitations section, without a heading, after the results and before the conclusion sentences.

We voluntarily did not use the same terminology as PRISMA because it may not be well understood by the readers (e.g., imprecision, inconsistency of effect or association and indirectness of the evidence are complex concepts). We kept the wording similar to that used by the authors of the systematic reviews, for example, when reporting the limitations in the discussion section. Another author (IB) read the entire modified abstract to determine whether the limitations section was written according to our specific guidance, that is, whether the limitations section 1) focused on the limitations previously identified in the report of the systematic review, 2) was no longer than 2 sentences, and 3) focused on limitations as described in the PRISMA for abstracts. When the second

author found that the limitation section was not written according our specific guidance, the limitations section was re-written and discussed to reach consensus.

Construction of abstracts with and without a limitations section

For each of the 30 selected abstracts, we obtained 1 abstract without a limitations section (i.e., the original abstract) and 1 with a limitations section (i.e., the original abstract plus the constructed limitations section). If the original abstract reported the limitations of the systematic review, this was deleted from the abstract without a limitations section and kept in the abstract with a limitations section. This situation occurred for 9 (30%) abstracts.

All abstracts were standardized, with the treatment name (s), authors and journal masked. The experimental treatment name was replaced by "intervention A". When the control treatment was an active treatment, "comparator B" replaced the treatment name. Acronyms and abbreviations were also deleted.

An example of an abstract with and without a limitations section is given in the Table 1 and all abstracts with a limitations section are available in the Additional file 1.

Participants

Eligible participants were the corresponding authors of clinical trials published between January 2010 and June

2013 and indexed in PubMed Core Clinical Journals, with an email address available on PubMed.

Potential participants were invited by e-mail to participate in an online survey on the interpretation of abstracts of systematic reviews. As an incentive, they were told that they would be entered into a draw along with all other participants for a chance to win an Apple iPad mini. If they agreed to participate, they logged onto a secure website and answered a screening question asking if they were a clinician; non-clinicians were excluded. Invitation emails were sent in waves until the planned number of clinicians had logged on and completed the assessment. A maximum of 2 reminders were sent to participants. The email invitation and details of the questionnaire are in the Additional file 1.

Randomization

A computer-generated randomization list was generated for allocating participants to abstracts with or without a limitations section in a 1:1 ratio. Allocation was concealed by use of a web-based randomization procedure whereby participants logged onto the system and were randomized to evaluate an abstract with or without a limitations section. Clinicians who logged onto the system but did not evaluate an abstract were excluded and the abstract was automatically allocated to another clinician.

Table 1 Example of abstract construction

Abstract without limitations section	Abstract with limitations section
<p>TITLE: Comparative effectiveness of intervention A and comparator B for treatment of advanced urothelial carcinoma.</p> <p>BACKGROUND: Intervention A is a standard treatment of metastatic urothelial carcinoma (UC), though comparator B is frequently substituted due to improved tolerability. Because comparative effectiveness in clinical outcomes of intervention A - versus comparator B chemotherapy is lacking, a meta-analysis was carried out.</p> <p>METHODS: PubMed was searched for articles published from 1966 to 2010. Eligible studies included prospective randomized trials evaluating intervention A - versus comparator B regimens in patients with metastatic UC. Individual patient data were not available and survival data were inconsistently reported. Therefore, the analysis focused on overall response (OR) and complete response (CR) rates. The Mantel-Haenszel method was used for combining trials and calculating pooled risk ratios (RRs).</p> <p>RESULTS: A total of 286 patients with metastatic UC from four randomized trials were included. Intervention A was associated with a significantly higher likelihood of achieving a CR [RR =3.54; 95% confidence interval (CI) 1.48-8.49; P =0.005] and OR (RR =1.34; 95% CI 1.04-1.71; P =0.02). Survival end points could not be adequately assessed due to inconsistent reporting among trials.</p> <p>CONCLUSIONS: Intervention A, as compared with comparator B, significantly increases the likelihood of both OR and CR in patients with metastatic UC. The impact of improved response proportions on survival end points could not be assessed.</p>	<p>TITLE: Comparative effectiveness of intervention A and comparator B for treatment of advanced urothelial carcinoma.</p> <p>BACKGROUND: Intervention A is a standard treatment of metastatic urothelial carcinoma (UC), though comparator B is frequently substituted due to improved tolerability. Because comparative effectiveness in clinical outcomes of intervention A - versus comparator B chemotherapy is lacking, a meta-analysis was carried out.</p> <p>METHODS: PubMed was searched for articles published from 1966 to 2010. Eligible studies included prospective randomized trials evaluating intervention A - versus comparator B regimens in patients with metastatic UC. Individual patient data were not available and survival data were inconsistently reported. Therefore, the analysis focused on overall response (OR) and complete response (CR) rates. The Mantel-Haenszel method was used for combining trials and calculating pooled risk ratios (RRs).</p> <p>RESULTS: A total of 286 patients with metastatic UC from four randomized trials were included. Intervention A was associated with a significantly higher likelihood of achieving a CR [RR =3.54; 95% confidence interval (CI) 1.48-8.49; P =0.005] and OR (RR =1.34; 95% CI 1.04-1.71; P =0.02). Survival end points could not be adequately assessed due to inconsistent reporting among trials.</p> <p>LIMITATIONS: This review is limited by the small sample sizes and methodological quality of the included studies. None of the included studies was blinded or placebo controlled; two studies closed early.</p> <p>Conclusions: Intervention A, as compared with comparator B, significantly increases the likelihood of both OR and CR in patients with metastatic UC. The impact of improved response proportions on survival end points could not be assessed.</p>

Blinding

Participants were blinded to the study's hypothesis. They were informed that they were participating in a study on the interpretation of abstracts of systematic reviews, but they were not aware that they were randomized to assess an abstract with or without a limitations section.

Outcomes measures

The primary outcome was readers' confidence in the results of the systematic review as stated in the abstract (i.e., *How confident are you in the results of the systematic review?*) assessed on a Likert scale, from 0 "not at all confident" to 10 "very confident". The secondary outcomes were the confidence in the validity of the conclusions (i.e., *How confident are you in the validity of the conclusion of this study?*), the beneficial effect of the experimental intervention (i.e., *How confident are you that intervention "A" could be of benefit to patients?*), the influence of the results on clinical practice (i.e., *How confident are you that the results of this study could influence your clinical practice?*) and the rigor of the systematic review (i.e., *Do you think that this systematic review was conducted rigorously?*) assessed on an 11-point Likert scale.

Sample size

Each participant read 1 abstract with or 1 abstract without a limitations section. The unit of analysis was the abstract. A sample of 266 participants was theoretically needed to be able to detect an effect size of 0.4 with the primary outcome (with power of 90% and alpha risk 5%). An effect size of 0.4 is equivalent to a decrease in the primary outcome of 1 point (the minimum expected difference between groups on a 0–10 scale) with an SD of 2.5. Theoretically, each abstract must be read the same number of times according to the randomization group. Knowing that each participant would read only one abstract (with or without a limitations section), we chose to include 300 participants. Each abstract with and without a limitations section was assessed 5 times.

Statistical analysis

Statistical analysis involved use of SAS v9.3 (SAS Inc., Cary, NC). All outcomes were quantitative; differences between groups were analyzed by a linear mixed model with a fixed factor (group) and random abstracts and abstract \times group interaction effects. A random-effects model allowed for taking into account 2 levels of clustering: by abstract (each abstract was assessed 5 times in each group) and interclustering (pairing between the abstracts used in the 2 arms of the trial). Inference was based on restricted maximum likelihood. For primary outcome and secondary outcomes, we estimated the difference between means (with 95% confidence intervals [95% CIs]) for abstracts

with and without a limitations section. $P < 0.05$ was considered statistically significant.

Ethical considerations

The institutional review board from the University of Paris Descartes approved the protocol No. CL178200001. The study is registered in ClinicalTrials.gov (no. NCT01848782).

Results

Characteristics of participants

Among the 4,807 potential participants who were invited by e-mail to participate in the survey between May 1 and June 30, 2013, 394 logged onto the study site; 89 were excluded because they were not clinicians and 5 did not complete the survey. From the 300 participants, 150 were randomized to the intervention group (i.e., abstracts with a limitations section), and 150 to the control group (i.e., abstracts without a limitations section). In total, 150 participants per group were included in the final analysis (Figure 1).

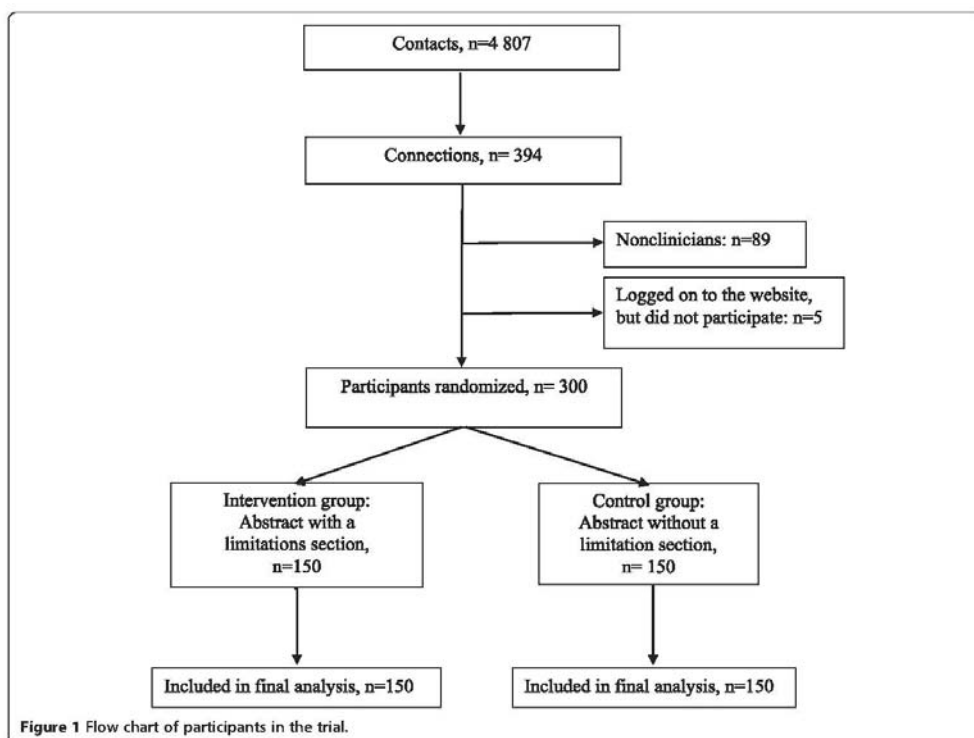
The median [Q1–Q3] participant age was 45 [range 38–54] years; 72% were male. Participants were mainly located in the European Union (49%) and in the United States (33%). Most had medical experience (74.7% had been clinicians for more than 10 years) and more than half (59%) regularly read reports of systematic reviews. More than half (53%) had been involved in a systematic review, but 46% had never peer-reviewed a systematic review for a biomedical journal (Table 2).

Characteristics of systematic reviews, abstracts and limitations sections

The general characteristics of the included systematic reviews and the quality of the reporting of the abstracts (according to the PRISMA statement for abstracts [12]) are in Table 3.

Overall, 23 abstracts were structured abstracts, and we created a specific heading for the limitations section, whereas 7 abstracts were not structured, and a sentence reporting the limitations was added before the conclusions. Some limitations were reported in 9 original abstracts. These limitations were deleted and rewritten according to the PRISMA guidelines. The quality assessment of the risk of bias in the selected systematic reviews was assessed with different tools. Authors of the included systematic reviews used the Cochrane or a modified Cochrane Collaboration tool in 10 abstracts (33%), the Jadad scale or modified Jadad scale in 5 (17%), the PEDro scale or modified PEDro scale in 3 (10%), and other tools in 4 (13%); the scale was not specified in 6 (20%).

The limitations sections we created focused on risk of bias in 22 abstracts (73%), heterogeneity in 13 (43%),



publication bias in 15 (50%), imprecision of results in 12 (40%), and indirectness of the evidence in 2 (7%) abstracts. The median [Q1-Q3] word count for limitations sections was 27 [23–31]. The median number of limits described in the limitations sections was 2 (range 1–4).

Clinicians' interpretation of abstracts

Readers' assessment of abstracts with and without a limitations section did not differ in the primary outcome – confidence in the results (mean [SD] =4.4 [2.3] and 4.6 [2.5], respectively; mean difference [95% CI] 0.2 [–0.4 to 0.7], $p = 0.5$). For the secondary outcomes, the assessment of abstracts with and without a limitations section did not differ for confidence in the validity of the conclusions (mean difference 0.07 [–0.5 to 0.6], $p = 0.8$); benefit of the experimental intervention to patients (mean difference 0.1 [–0.4 to 0.7], $p = 0.6$); influence of the results on clinical practice (mean difference –0.08 [–0.6 to 0.5], $p = 0.8$) and rigor of the systematic review (mean difference –0.4 [–1.0 to 0.2], $p = 0.2$) (Table 4).

Discussion

This study evaluated, in a randomized controlled trial, the impact of adding a limitations section to an abstract on the interpretation of abstracts of systematic reviews. This randomized controlled trial involved a large international sample of clinicians and a sample of “real life” abstracts of systematic reviews (i.e., abstracts of published systematic reviews indexed in DARE). Despite the selection of a sample of systematic reviews of good quality, the mean confidence of readers was low, and adding a limitations section had no impact on the interpretation of abstract results by expert-readers.

Because abstracts are the first and sometimes the only source of information for readers, editors are attentive to their quality and their capacity to provide all the necessary and important information on the research performed. In the 1960s, abstracts were usually reported on the last page of research articles and were moved to the beginning of research articles [18]. Since then, many editorial policies have been implemented to try to improve the content and the format of abstracts. These policies

Table 2 Characteristics of participants

		Abstract without limitations n (%)	Abstract with limitations n (%)	Total n (%)
		N = 150	N = 150	N = 300
Qualification	MD	78 (52.0)	78 (52.0)	156 (52.0)
	PhD	58 (38.7)	64 (42.7)	122 (40.7)
	Other	14 (9.3)	8 (5.3)	22 (7.3)
Clinical experience	<5 years	6 (4.0)	8 (5.3)	14 (4.7)
	5 to 10 years	34 (22.7)	28 (18.7)	62 (20.7)
	>10 years	110 (73.3)	114 (76.6)	224 (74.7)
Reading reports of systematic reviews	Rarely or sometimes	61 (40.7)	62 (41.6)	123 (41.1)
	Regularly	89 (59.3)	87 (58.4)	176 (58.9)
No. of randomized trials involved in	0	16 (10.7)	11 (7.3)	27 (9.0)
	1-3	48 (32.0)	48 (32.0)	96 (32.0)
	4-9	45 (30)	54 (36.0)	99 (33.0)
	>10	41 (27.3)	37 (24.7)	78 (26.0)
Authored at least one systematic review		77 (51.3)	82 (55.0)	159 (53.2)
No. of systematic reviews peer-reviewed	0	75 (50.0)	64 (43.0)	139 (46.5)
	1-3	55 (36.7)	69 (46.3)	124 (41.5)
	>3	20 (13.3)	16 (10.7)	36 (12.0)
Training in clinical epidemiology		101 (67.3)	76 (50.7)	177 (59.0)
Training in methods of randomized trials		86 (57.3)	71 (47.3)	157 (52.3)

have involved the development and implementation of structured abstracts [18,19], reporting guidelines such as the CONSORT for abstracts of randomized controlled trials and the PRISMA statement for abstracts of systematic reviews [12,20]. Such policies can improve the quality of reporting of abstracts [21] and should in theory improve the interpretation by readers.

However, despite these initiatives, the quality of reporting of abstracts remains questionable [9,10,22-24]. A recent study showed that despite systematic reviews including primary studies with high risk of bias, just over half included a risk of bias assessment in the interpretation of results in the abstract [13]. Consequently, adding a limitations section could be useful to enhance readers' awareness and improve their interpretation. However, a limitations section in the abstract is recommended by only a few journals and for systematic reviews in the PRISMA statement for abstracts [12]. For example, *Annals of Internal Medicine* has required authors to include a limitations section in the abstract of scientific articles since 2004 [14].

To our knowledge, the impact of adding a limitations section in abstracts of systematic reviews has never been evaluated. Previous studies have evaluated the impact of different reporting on the interpretation of the study by readers. These studies mainly involved use of a single abstract of a fictional trial. For example, industry sponsorship can negatively influence the perception of the methodological quality of a study and the willingness

to believe the study findings [25]. Similarly, interpretation of study results is affected by the reporting of outcomes as absolute risk, relative risk or number needed to treat [26,27].

Our results did not show any impact of the abstract limitations section on expert-readers' interpretation. Furthermore, our results highlight that confidence in results was low in both arms. The high level of expertise of the participants could explain these results. In fact, half of the clinicians included in this randomized controlled trial had some experience in the conduct and reviewing of systematic reviews. This level of expertise could increase their awareness of the common limitations of systematic reviews such as the risk of publication bias or the limited quality of primary studies. Furthermore, the limitations section reported factual information and in a neutral form, and the conclusion of the systematic reviews' abstract was not modified. Also, assessing the confidence in the results of a systematic review is complex because it depends both on how the systematic review is conducted and the quality of the primary studies included.

Our study has some limitations. First, the readers did not access the full-text article to fully appraise the study results; they only assessed an abstract with or without a limitations section. However abstracts of systematic reviews are very important, because some readers cannot access full-text articles because of the fee requirement,

Table 3 Characteristics of systematic reviews included

Characteristics		N = 30
Field of research, n (%)	Medicine	27 (90)
	Surgery	3 (10)
Impact factor, median; [Q1-Q3]; (min-max)		4.60; [3.18-7.51]; (0.00-18.04)
No. of trials included, median; [Q1-Q3]; (min-max)		13; [6-23]; (4-54)
Sample size, median; [Q1-Q3]; (min-max)		2872; [801-11005]; (257-782460)
Experimental treatment, n (%)	Drugs	16 (53)
	Complex intervention	11 (37)
	Surgery	2 (7)
	Device	1 (3)
Meta-analysis included in the systematic review, n (%)		26 (87)
Funding source, n (%)	Non-profit	11 (37)
	For-profit	2 (7)
	None	10 (33)
	Not specified	7 (23)
Content of the original abstract, n (%)	Eligibility criteria	17 (57%)
	Key data bases and date of search	9 (30%)
	Method to assess the risk of bias	6 (20%)
	No. and type of the studies included in the systematic review	22 (73%)
	Summary measure and confidence interval for the main outcome results*	20/26 (77%)
	Strength and limitation	13 (43%)

*This result applies on the 26 systematic reviews that included a meta-analysis.

low Internet download capacity, or the full-text article being available only in a language not understood by the reader. Second, the participants were corresponding authors of articles of randomized trials and systematic reviews, who may be considered "reader-experts" and not representative of all readers of systematic reviews. Consequently, we cannot exclude that a limitations section may be useful for a less expert readership. Finally, we explored only the impact of a limitations section added to abstracts reporting a systematic review and we cannot

extrapolate our findings to limitations sections in abstracts reporting other types of studies such as randomized controlled trials and observational studies.

However, this study has important implications. At this stage, we cannot make any recommendations for practice and we should probably not change guidelines and editorial policies related to the reporting of a limitations section in abstracts of systematic reviews. However, this study highlights an important topic for future research. First, because our study is the first study on

Table 4 Results of primary and secondary outcomes

	Abstract without limitations Mean (SD) n = 150	With Limitations Mean (SD) n = 150	Mean difference (95% CI)	p-value
How confident are you in the results of this study (0-10)?	4.6 (2.5)	4.4 (2.3)	0.2 (-0.4 to 0.7)	p = 0.5
How confident are you in the validity of the conclusion of this study (0-10)?	4.1 (2.5)	4.0 (2.3)	0.07 (-0.5 to 0.6)	p = 0.8
How confident are you that the intervention A could be benefit to patients (0-10)?	4.4 (2.6)	4.3 (2.3)	0.1 (-0.4 to 0.7)	p = 0.6
How confident are you that the results of this study could influence your clinical practice (0-10)?	3.8 (2.6)	3.8 (2.3)	-0.08 (-0.6 to 0.5)	p = 0.8
Do you think that this systematic review was conducted rigorously (0-10)?	4.1 (2.7)	4.4 (2.6)	-0.4 (-1.0 to 0.2)	p = 0.2

95% CI, 95% confidence interval.

this topic, the trial should be replicated, and other trials including a less expert readership or with different background (e.g. authors of "clinical practice guidelines") should be performed. Second, qualitative studies would probably be useful to help define how limitations sections should be reported to have a real impact on readers. Third, we recommend exploring the impact of a limitations section in abstracts of other study designs such as randomized controlled trials and observational studies. Overall, more research is needed on the interpretation of research results from abstracts because abstracts are widely disseminated.

Conclusions

In conclusion, adding a limitations section in abstracts of systematic review may not affect expert-readers' interpretation of abstract results and conclusions. Future studies are needed to confirm these results and explore the impact of a limitations section on a less expert panel of participants.

Additional file

Additional file 1: The 30 Abstracts with limitations section. The invitation e-mail for the participants. The survey.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived, designed and experiments: AY, IB, SH, PR. Performed the experiments: AY, IB. Analyzed the data: GB. Wrote the first draft: AY, IB. Contributed to the writing of the manuscript: AY, IB, SH, GB, PR. All authors read and approved the final manuscript.

Authors' information

On-line survey design and data management: Isabelle Pane, senior computer engineer and data manager, and Joan Denis, junior computer engineer and data manager, Centre de recherche Épidémiologies et Biostatistiques, INSERM U1153, Paris, France.

Acknowledgements

We thank all clinicians who participated to the study.

Financial support

AY was funded by a grant from the Fondation pour la Recherche Médicale.

Transparency declaration

AY affirms that this 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. AY and IB had access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Author details

¹Centre de Recherche Épidémiologies et Statistiques INSERM U1153, Paris, France. ²Centre of Clinical Epidemiology, Assistance Publique des Hôpitaux de Paris, Hôtel Dieu Hospital, Place du Parvis Notre-Dame, 75181 Paris, Cedex 4, France. ³Faculté de Médecine, French Cochrane Center, Paris Descartes University, Sorbonne Paris Cité, Paris, France. ⁴Department of Anesthesiology and Critical Care, Beaujon University Hospital, Clichy, France. ⁵Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. ⁶Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford OX2 6UD, UK.

Received: 29 July 2014 Accepted: 13 November 2014
Published: 24 November 2014

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn't. *BMJ* 1996, **312**:71-72.
2. 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**:e1000326.
3. Cook DJ, Mulrow CD, Haynes RB: Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997, **126**:376-380.
4. Bero LA, Jadad AR: How consumers and policymakers can use systematic reviews for decision making. *Ann Intern Med* 1997, **127**:37-42.
5. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM: Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007, **7**:10.
6. Saint S, Christakis DA, Saha S, Elmore JG, Welsh DE, Baker P, Koepsell TD: Journal reading habits of internists. *J Gen Intern Med* 2000, **15**:881-884.
7. Scherer RW, Langenberg P, von Elm E: Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007, (2):MR000005.
8. The PLoS Medicine Editors: The impact of open access upon public health. *PLoS Med* 2006, **3**:e252. doi:10.1371/journal.pmed.0030252.
9. Hopewell S, Clarke M, Askie L: Reporting of trials presented in conference abstracts needs to be improved. *J Clin Epidemiol* 2006, **59**:681-684.
10. Beller EM, Glasziou PP, Hopewell S, Altman DG: Reporting of effect direction and size in abstracts of systematic reviews. *JAMA* 2011, **306**:1981-1982.
11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009, **339**:b2700.
12. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, Gøtzsche PC, Lasserson T, Tovey D: PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013, **10**:e1001419.
13. 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**:e003342.
14. The Editors: Addressing the limitations of structured abstracts. *Ann Intern Med* 2004, **140**:480-481.
15. Schulz KF, Altman DG, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010, **63**:834-840.
16. Guyatt GH: GRADE: what is "quality of evidence" and why is it important to clinicians? *BMJ* 2008, **336**:995-998.
17. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ: GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008, **336**:924.
18. Soffler A: Abstracts of clinical investigations. A new and standardized format. *Chest* 1987, **92**:389-390.
19. Huth EJ: Structured abstracts for papers reporting clinical trials. *Ann Intern Med* 1987, **106**:626.
20. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF: CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008, **5**:e20.
21. 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**:e1178.
22. Ioannidis JP, Lau J: Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001, **285**:437-443.
23. Yank V, Rennie D, Bero LA: Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ* 2007, **335**:1202-1205.
24. 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**:2058-2064.

25. Kesselheim AS, Robertson CT, Myers JA, Rose SL, Gillet V, Ross KM, Glynn RJ, Joffe S, Avorn J: **A randomized study of How physicians interpret research funding disclosures.** *N Engl J Med* 2012, **367**:1119–1127.
26. Bucher HC, Weinbacher M, Gyr K: **Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration.** *BMJ* 1994, **309**:761–764.
27. Farrow L, Taylor WC, Arnold RM: **Absolutely relative: how research results are summarized can affect treatment decisions.** *Am J Med* 1992, **92**:121–124.

doi:10.1186/1471-2288-14-123

Cite this article as: Yavchitz et al: **Impact of adding a limitations section to abstracts of systematic reviews on readers' interpretation: a randomized controlled trial.** *BMC Medical Research Methodology* 2014 **14**:123.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



4. Classification des spin dans les revues systématiques et les méta-analyses

4.1. Justification de l'étude

Les revues systématiques et les méta-analyses ont un rôle majeur pour la pratique de « l'Evidence Based Medicine »[4,135]. Une présentation et une interprétation non biaisées de ces études sont essentielles pour implémenter leurs résultats en pratique clinique. Par ailleurs, la présence de *spin* dans les publications scientifiques peut avoir des conséquences importantes sur l'interprétation et la diffusion des résultats de la recherche[136,137]. Aussi, il n'existe pas à notre connaissance de classification des *spin* dans les revues systématiques et les méta-analyses.

4.2. Objectifs

Les objectifs de cette étude sont : 1) développer une classification des *spin* dans les revues systématiques et les méta-analyses, et 2) classer les *spin* du titre et du résumé selon leur sévérité.

4.3. Méthodes

4.3.1. Définitions

Nous avons défini les *spin* comme une manière de rapporter l'étude, délibérément ou non, de façon à convaincre les lecteurs que l'intervention est plus bénéfique que ce que montrent les résultats de l'étude ; c'est-à-dire à surestimer l'effet bénéfique ou à sous-estimer les effets secondaires de l'intervention évaluée. Nous n'avons pas pris en compte les *spin* spécifiques qui sous-estiment l'effet de l'intervention en termes d'efficacité ou de sécurité.

Tous les auteurs de l'étude sont experts dans le domaine de la réalisation et de l'écriture des rapports de revues systématiques et méta-analyses ; tous ont participé aux différentes étapes de l'étude.

4.3.2. Développement de la classification des *spin* dans les revues systématiques et les méta-analyses

Pour identifier et classer les différents *spin* dans les revues systématiques et les méta-analyses nous avons procédé en quatre étapes.

Etape 1 : Nous avons effectué une revue de littérature sur les problèmes de conduite, d'interprétation ou de présentation des revues systématiques et méta-analyses évaluant des interventions thérapeutiques pour identifier les potentiels *spin*. Nous avons cherché les sources suivantes : 1) le Cochrane Handbook, 2) le « Methodological standards for conduct of

Cochrane Intervention Review » (MECIR) version 2, 3) les bases de données MEDLINE et Cochrane Methodology Register et 4) les articles princeps et les recommandations sur les revues systématiques et les méta-analyses. A partir de cette revue de littérature nous avons développé une classification préliminaire.

Phase 2 : Nous avons classé les potentiels *spin* identifiés lors de la première étape selon leur localisation : dans le texte intégral ou dans le résumé de l'article afin d'obtenir un classement préliminaire.

Etape 3 : Nous avons organisé des entretiens individuels avec chacun des auteurs de la présente étude. Lors de ces entretiens ouverts nous avons discuté chaque item de la classification préliminaire, puis les auteurs avaient l'opportunité d'exprimer des éclaircissements, opinions, justification, définition ou de nouvelles idées.

Etape 4 : Nous avons finalisé la classification des *spin* dans les revues systématiques et méta-analyses en incorporant les commentaires des auteurs jusqu'à l'obtention d'un consensus.

4.3.3. Classement des spin dans les résumés des revues systématiques et méta-analyses

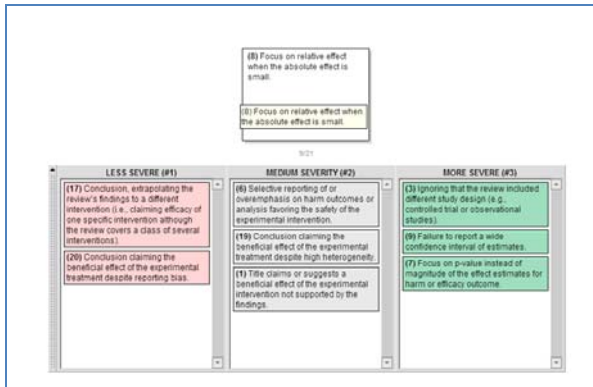
Le classement a porté uniquement sur les *spin* des résumés. Pour classer les *spin* nous avons utilisé la méthode du Q-sort, développé en 1935 par W. Stephenson. Cette méthode associe une approche quantitative et une approche qualitative pour étudier la subjectivité ou les différents points de vue[138–141]. Nous avons réalisé l'enquête en ligne grâce au logiciel

FlashQ version 1.0[142]. L'enquête s'est déroulée en trois étapes. Chaque item correspond à un *spin*. (1) Les participants devaient lire la liste des items et les ranger dans une des trois piles selon leur sévérité potentielle « moins sévère », « sévérité modérée », « plus sévère ». (2) Les participants devaient ensuite placer chaque item sur une grille de score allant de -3 : le moins sévère, à +3 : le plus sévère. (3) Les participants avaient la possibilité de relire les items et de modifier leur emplacement sur la grille de score autant de fois que nécessaire.

Les étapes de l'enquête sont représentées sur la figure 3.

Nous avons invité à participer à l'enquête tous les coordinateurs des groupes Cochrane, tous les membres du « Methodology Review Group » et du « Bias Methods Group » ainsi que tous les éditeurs des groupes les plus actifs. Les participants étaient invités par mail.

Les *spin* ont ensuite été classés selon la moyenne des scores donnés par les participants.



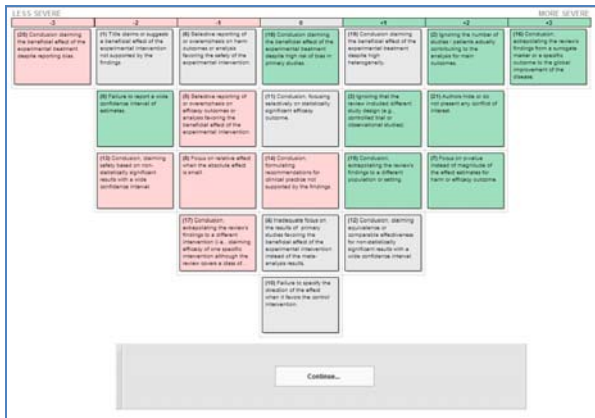
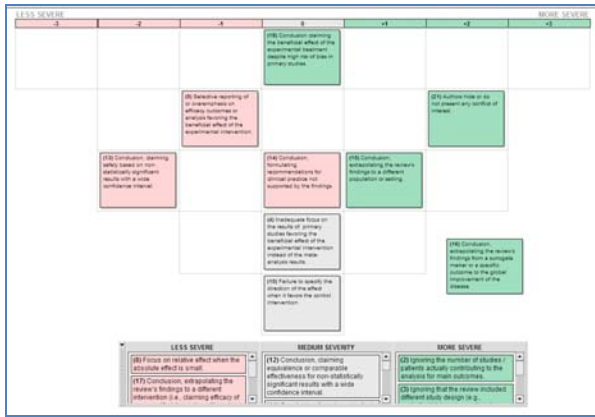


Figure 3: Les étapes de l'enquête en ligne

4.4. Résultats

4.4.1. Identification des spin dans les revues systématiques et les méta-analyses

A partir de notre revue de littérature nous avons identifié 39 *spin*, 28 pour le texte intégral et onze pour le résumé de l'article. Parmi les *spin* identifiés, 13 (33%) sont spécifiques des revues systématiques et méta-analyses, et 26 (67%) pourraient être retrouvés dans d'autres

types d'études. Par exemple le *spin* qui consiste en l'extrapolation des résultats d'un critère biologique à l'amélioration globale de la maladie peut être retrouvé aussi bien dans une revue systématique que dans un essai contrôlé randomisé. Par contre le *spin* qui consiste à focaliser l'attention sur les résultats des études incluses dans la revue systématique à la place des résultats de la méta-analyse est spécifique à ce plan d'étude. Les 13 *spin* spécifiques des revues systématiques et méta-analyses sont surlignés de gris dans le tableau 7.

4.4.2. Classification des spin dans les revues systématiques et les méta-analyses

Nous avons classé les *spin* en trois grandes catégories selon le type de stratégie utilisée par les auteurs de la revue. Les trois catégories sont :

- 1) **Une présentation inadéquate** de la revue que nous avons définie comme une présentation incomplète ou inappropriée de la méthode, des analyses réalisées, des résultats ou de toute information importante pour la compréhension de la revue et qui pourrait fausser l'interprétation du lecteur.
- 2) **Une interprétation inadéquate** des résultats de l'étude, que nous avons définie comme une interprétation par les auteurs pouvant induire le lecteur en erreur.
- 3) **Une extrapolation inadéquate** définie comme une généralisation inappropriée des résultats de l'étude.

La classification des *spin* est présentée dans le tableau 7.

4.4.3. Résultats de l'enquête sur la sévérité des spin dans les résumés des revues systématiques et méta-analyses

L'enquête s'est déroulée du 27 février au 10 avril 2015. Parmi les 476 experts invités à participer à l'étude, 122 (26%) ont complété l'enquête en ligne. Les participants avaient un haut niveau d'expertise dans la conduite ou la relecture des revues systématiques ; 90% (n=107) avaient déjà participé à la réalisation d'une revue systématique et 42% (n=51) étaient affiliés à l'un des groupes méthodes de la collaboration Cochrane.

Les scores moyens allaient de 1,83 (1,18) à -0,96 (1,34) et huit stratégies de *spin* avaient une moyenne positive. Les trois stratégies de *spin* classées par les participants comme étant les plus sévères étaient : 1) Conclure en formulant des recommandations pour la pratique non supportées par les résultats, 2) Le titre déclare ou suggère un effet bénéfique de l'intervention non supporté par les résultats, 3) Présentation sélective ou mise en avant des critères d'efficacité dont le résultat est en faveur de l'effet bénéfique de l'intervention. Le classement des *spin* dans le titre et le résumé classé par ordre croissant de sévérité est présenté dans le tableau 8.

Tableau 7 : Classification des *spin* dans les revues systématiques et les méta-analyses.

Stratégies	<i>Spin</i> dans le texte intégral	<i>Spin</i> dans le résumé
Présentation inadéquate	<ol style="list-style-type: none"> 1) Ne pas rapporter un écart au protocole qui pourrait modifier l'interprétation des résultats 2) Présentation sélective ou mise en avant inappropriée des critères d'efficacité favorisant l'effet bénéfique de l'intervention expérimentale (critères secondaire, analyse en sous-groupe) 3) Présentation sélective ou mise en avant inappropriée des critères de sécurité favorisant la sûreté de l'intervention expérimentale (critères secondaires, analyse en sous-groupe) 4) Présentation sélective des analyses favorisant l'effet bénéfique de l'intervention expérimentale (modèle à effet fixe ou à effet aléatoire) 5) Mise en avant inadéquate des résultats des études primaires favorisant l'effet bénéfique de l'intervention expérimentale à la place des résultats de la méta-analyse 6) Changer l'échelle du <i>forest plot</i> pour magnifier les résultats 7) Présentation inadéquate ou absente des limites des revues systématiques 8) Citation sélective des articles favorisant l'effet bénéfique de l'intervention expérimentale 9) Cacher ou ne pas déclarer les conflits d'intérêts 10) Conclusion uniquement basée sur les critères d'efficacité statistiquement significatifs 	<ol style="list-style-type: none"> 1) Présentation sélective ou mise en avant inappropriée dans le résumé des critères d'efficacité favorisant l'effet bénéfique de l'intervention expérimentale 2) Présentation sélective ou mise en avant inappropriée dans le résumé des critères de sécurité favorisant la sûreté de l'intervention expérimentale 3) Ne pas rapporter dans le résumé le nombre d'études ou de patients participant réellement à l'analyse des critères principaux 4) Ne pas rapporter dans le résumé un large intervalle de confiance de l'effet estimé 5) Ne pas rapporter dans le résumé la direction de l'effet estimé quand celui est en faveur de l'intervention contrôle. 6) Cacher ou ne pas déclarer les conflits d'intérêts 7) Mise en avant inadéquate des résultats des études primaires favorisant l'effet bénéfique de l'intervention expérimentale à la place des résultats de la méta-analyse 8) Conclusion uniquement basée sur les critères d'efficacité statistiquement significatifs

<p>Interprétation inadéquate</p>	<p>11) Le titre suggère ou déclare à tort l'efficacité de l'intervention expérimentale</p> <p>12) Résultats non statistiquement significatifs inadéquatement interprétés comme un manque d'effet ou comme une équivalence d'effet entre les deux bras d'études.</p> <p>13) Résultats non statistiquement significatifs inadéquatement interprétés comme une preuve de sûreté pour un critère de sécurité.</p> <p>14) Mise en avant inadéquate de la valeur du p au lieu de l'amplitude de l'effet estimé.</p> <p>15) Mise en avant du risque relatif quand le risque absolu est petit.</p> <p>16) Pas de prise en compte ou prise en compte insuffisante de l'hétérogénéité dans l'interprétation des résultats</p> <p>17) Pas de prise en compte ou prise en compte insuffisante du risque de biais des études primaires dans l'interprétation des résultats</p> <p>18) Pas de prise en compte ou prise en compte insuffisante du biais de publication dans l'interprétation des résultats</p> <p>19) Interprétation erronée des articles cités, favorisant à tort l'effet bénéfique du traitement expérimental</p> <p>20) Conclusion à tort à l'équivalence ou l'efficacité comparable pour des résultats non statistiquement significatifs avec un large intervalle de confiance</p> <p>21) Conclure à tort sur la sécurité en s'appuyant sur des résultats non statistiquement significatifs avec un large intervalle de confiance</p> <p>22) Conclusion ignorant à tort un risque élevé de biais, l'hétérogénéité ou le biais de publication dans l'interprétation des résultats.</p>	<p>9) Le titre suggère ou déclare à tort l'efficacité de l'intervention expérimentale</p> <p>10) Ne pas rapporter dans le résumé que la revue inclut des études avec différents plans (ERC et études observationnelles)</p> <p>11) Mise en avant inadéquate de la</p> <p>12) valeur du p au lieu de l'amplitude de l'effet estimé</p> <p>13) Mise en avant du risque relatif quand le risque absolu est petit</p> <p>14) Conclure à tort à l'équivalence ou l'efficacité comparable pour des résultats non statistiquement significatifs avec un large intervalle de confiance</p> <p>15) Conclure à tort sur la sécurité en s'appuyant sur des résultats non statistiquement significatifs avec un large intervalle de confiance</p> <p>16) Conclure à tort à l'efficacité de l'intervention expérimentale malgré un risque de biais élevé dans les études primaires.</p> <p>17) Conclure à tort à l'efficacité de l'intervention expérimentale malgré l'hétérogénéité</p> <p>18) Conclure à tort à l'efficacité de l'intervention expérimentale malgré un biais de publication</p>
---	---	---

<p>Misleading Extrapolation</p>	<p>23) Extrapolation inadéquate des résultats d'un critère intermédiaire ou spécifique vers l'amélioration globale de la maladie.</p> <p>24) Extrapolation inadéquate des résultats de la revue à une population plus large, un cadre plus large ou un ensemble plus large d'interventions</p> <p>25) Conclure à tort en extrapolant les résultats de la revue à une population différente ou à une localisation différente</p> <p>26) Conclure à tort en extrapolant les résultats de la revue à une intervention différente</p> <p>27) Conclure à tort en extrapolant les résultats d'un critère intermédiaire ou spécifique à l'évolution globale de la maladie.</p> <p>28) La conclusion émet à tort des recommandations pour la pratique clinique</p>	<p>19) Dans le résumé, extrapolation inadéquate des résultats de la revue à une population ou à une localisation différente</p> <p>20) Dans le résumé, extrapolation inadéquate des résultats de la revue à une intervention différente</p> <p>21) Dans le résumé, extrapolation inadéquate des résultats d'un critère intermédiaire ou spécifique vers l'amélioration globale de la maladie.</p> <p>22) La conclusion émet à tort des recommandations pour la pratique clinique</p>
--	--	--

Tableau 8: Classement des *spin* selon leur sévérité

Stratégie de <i>spin</i>	Moyenne (DS) N=122	Rang
La conclusion émet à tort des recommandations pour la pratique clinique	1.83 (1.18)	1
Le titre suggère ou déclare à tort l'efficacité de l'intervention expérimentale	1.43 (1.49)	2
Présentation sélective ou mise en avant des critères d'efficacité dont le résultat est en faveur de l'effet bénéfique de l'intervention	0.75 (1.20)	3
Conclure en déclarant qu'une intervention est sûre lorsque le résultat de l'analyse est statistiquement non significatif et l'intervalle de confiance large	0.46 (1.18)	4
Conclure à l'efficacité du traitement sans tenir compte du risque élevé de biais dans les études incluses.	0.43 (1.32)	5
Présentation sélective ou mise en avant des critères de sécurité en faveur de la sûreté de l'intervention	0.31 (1.45)	6
Conclure en extrapolant les résultats de la revue à une intervention différente	0.19 (1.27)	7
Conclure en extrapolant les résultats des critères intermédiaires ou spécifiques à une amélioration globale de la maladie	0.16 (1.42)	8
Conclure à l'efficacité du traitement sans tenir compte du biais de publication.	-0.12 (1.33)	9
Cacher ou ne pas déclarer les conflits d'intérêts	-0.16 (1.78)	10
Conclure en s'appuyant uniquement sur les critères d'efficacité statistiquement significatifs	-0.17 (1.54)	11
Conclure en déclarant que les interventions sont équivalentes lorsque le résultat de l'analyse est statistiquement non significatif et l'intervalle de confiance large	-0.19 (1.40)	12
Dans le résumé, ne pas préciser la direction de l'effet quand il est en faveur du groupe contrôle	-0.2 (1.62)	13
Ne pas rapporter l'intervalle de confiance des estimations	-0.29 (1.28)	14
Conclure en extrapolant les résultats de la revue à une population ou à une localisation différente	-0.43 (1.43)	15
Ne pas préciser que différents types d'essais ont été inclus	-0.44 (1.40)	16
Conclure à l'efficacité du traitement sans tenir compte de l'hétérogénéité	-0.51 (1.24)	17
Porter l'attention sur les résultats des études primaires favorisant l'effet bénéfique de l'intervention à la place des résultats de la méta-analyse	-0.63 (1.55)	18
Ne pas rapporter le nombre d'études et le nombre de patients ayant contribué à l'analyse du critère principale.	-0.69 (1.39)	19
Porter l'attention sur la valeur du p plutôt que sur l'amplitude de l'effet estimé	-0.78 (1.20)	20
Porter l'attention sur l'efficacité relative quand l'efficacité absolue est petite	-0.96 (1.34)	21

4.5. Discussion

Notre étude a permis d'élaborer une classification des *spin* dans les revues systématiques et les méta-analyses et de classer les *spin* des résumés selon leur sévérité perçue par des participants experts. Nous avons identifié 28 *spin* dans le texte intégral et 21 dans le résumé. Notre enquête a été réalisée auprès d'un large panel de spécialistes dans le domaine des revues systématiques et des méta-analyses, tous sont membres de la collaboration Cochrane. Les *spin* identifiés comme les plus sévères sont 1) La conclusion émet à tort des recommandations pour la pratique clinique, 2) Le titre suggère ou déclare à tort l'efficacité de l'intervention expérimentale et 3) Présentation sélective ou mise en avant des critères d'efficacité dont le résultat est en faveur de l'effet bénéfique de l'intervention.

Nous avons choisi de mener l'enquête sur la sévérité des *spin* uniquement pour les *spin* des résumés pour plusieurs raisons. 1) La diffusion des résultats de la recherche provient essentiellement des informations contenues dans les résumés des articles. 2) Le résumé est la partie de l'article la plus facilement accessible, elle est gratuite et indexée dans les bases de données comme MEDLINE. 3) Des études ont montré que les lecteurs évaluent les recherches principalement à la lecture du résumé[24,143].

Le classement des *spin* selon leur sévérité nous a permis de montrer que les *spin* dans le titre et la conclusion du résumé sont les plus à risque d'induire en erreur les lecteurs. En effet, ces deux parties du résumé sont celles qui délivrent un message rapide et facile à retenir. Elles devraient donc faire l'objet d'une attention particulière de la part des auteurs et des éditeurs.

Notre étude présente plusieurs limites. Premièrement, le classement des *spin* reflète la sévérité perçue par des experts dans le domaine des revues systématiques et des méta-analyses. Nous

n'avons pas étudié l'impact réel des *spin* sur l'interprétation des résultats par les lecteurs. Deuxièmement, nous avons étudié uniquement les *spin* dans les revues systématiques et les méta-analyses évaluant des interventions de santé chez l'homme. Les stratégies de *spin* pourraient être différentes dans les revues systématiques d'études diagnostiques ou de pronostiques par exemple. Enfin, notre classification n'est pas exhaustive. En effet il existe probablement d'autres stratégies de *spin* que nous n'avons pas encore identifiées. Néanmoins nous avons défini les trois grands types de stratégies de *spin*, ainsi de nouvelles stratégies de *spin* pourraient facilement être intégrées à cette classification.

En sensibilisant les auteurs, les éditeurs et les relecteurs, notre classification pourrait avoir un rôle important dans la prévention et la détection des *spin* dans les revues systématiques et les méta-analyses. Nous supposons que l'introduction de *spin* dans les rapports d'étude par les auteurs n'est pas toujours délibérée. Les *spin* peuvent résulter de l'ignorance ou du mimétisme des jeunes auteurs envers les plus anciens. Aussi les éditeurs et les relecteurs doivent avoir un rôle de garde-fou. Le classement des *spin* selon leur sévérité permet de hiérarchiser les *spin* les plus sévères et donc ceux qui doivent être évités à tout prix.

L'élaboration de la classification des *spin* dans les revues systématiques et les méta-analyses et leur classement selon leur sévérité est une première étape. De futurs travaux de recherche sont nécessaires pour déterminer la fréquence des *spin* dans les revues systématiques et les méta-analyses ainsi que leur impact sur l'interprétation des résultats par les lecteurs.

4.6. Article

Yavchitz A, Ravaud P, Moher D, Altman D, Lasserson T, Hrobjartsson A, Boutron I

Classification Scheme and Ranking of Spin in Systematic Reviews and Meta-Analyses

Article en cours

Classification Scheme and Ranking of Spin in Systematic Reviews and Meta-analyses

Amélie Yavchitz (1, 2, 3), Philippe Ravaud (1, 2, 3, 4), Doug Altman (5), David Moher (6,7), Asbjørn Hrobjartsson (8), Toby Lasserson (9), Isabelle Boutron (1, 2, 3)

- (1) Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153)
Inserm / Université Paris Descartes
- (2) Centre d'Épidémiologie Clinique, AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Hôtel Dieu, Paris, France
- (3) French Cochrane Center, Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, Paris, France
- (4) Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA
- (5) Centre for Statistics in Medicine, University of Oxford, Oxford, UK
- (6) Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada
- (7) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada
- (8) Nordic Cochrane Centre, Rigshospitalet Dept 7811, Copenhagen, Denmark
- (9) Cochrane Editorial Unit, London, United Kingdom

Corresponding author:

Dr Amélie Yavchitz

Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu

1 Place du Parvis Notre-Dame,

75181 PARIS Cedex 4

Tel 33 (0) 1 42 34 78 33 / 33 (0) 6 22 08 66 48

Fax 33 (0) 1 42 34 87 90

amelie.yavchitz@bjn.aphp.fr

ayavchitz@gmail.com

Word count abstract: 192

Word count full text: 2653

ABSTRACT

objectives

We aimed to 1) develop a classification of spin (i.e., misleading presentation or interpretation of study results) in systematic reviews and 2) rank spin in abstracts according to their severity (i.e., the likelihood of distorting readers' interpretation of the results).

STUDY DESIGN

We identified spin in systematic reviews and meta-analyses of therapeutic interventions by a literature review, then developed a classification scheme of spin in a four-phase consensus process. We ranked spin in abstracts according to their severity using a Q-sort method with experts from the Cochrane Collaboration.

RESULTS

We identified 39 types of spin, 28 for the main text and 21 for the abstract; 13 were specific to the systematic review and meta-analysis design. Spin types were classified in 3 categories: (1) misleading reporting, (2) misleading interpretation and (3) inadequate extrapolation. Spin ranked to be the most severe by 122 the experts who participated in the survey were spin in title and spin in conclusion such as recommendation not supported by findings.

CONCLUSION

Our classification could help authors avoid spin in reporting systematic review findings and editors and reviewers track spin to reduce the misinterpretation of review findings by readers.

Key words: Spin; Systematic reviews; Meta-analysis; Interpretation bias; Classification; Interpretation; Bias; Abstract; Misinterpretation; Distortion

What is new?

- We developed a classification of spin for the main text and abstract of systematic reviews. This classification contains 28 spin in the main text and 21 in the abstracts; 13 were specific to the systematic review and meta-analysis design.
- This spin classification could help editors and reviewers track spin and improve the interpretation of review findings and also help authors avoid spin when reporting a systematic review.
- The ranking of spin in abstract according to their severity highlighted that the most severe spin were spin in title, spin in the conclusion and selective reporting that favor beneficial effect or safety of the experimental treatment

BACKGROUND

Clinical and health-policy decision-making should be based on the best evidence. Systematic reviews and meta-analyses are the cornerstones of therapeutic evaluation because they summarize all available evidence. Clinicians, decision makers and researchers often refer to systematic reviews and meta-analyses to keep up to date with the medical literature, develop clinical practice guidelines and sometimes plan future research[1,2]. An accurate presentation and interpretation of results of systematic reviews is therefore essential for disseminating the results and applying them to clinical practice. However, authors usually have broad latitude in writing their reports of scientific results[3], with the possibility of adding “spin” (i.e., distorted presentation and interpretation of the results)[4].

The presence of spin in scientific reports was systematically studied for the first time in terms of the randomized controlled trial (RCT) design[4]. Spin was found in about 60% of abstract conclusions in RCT reports and their presence affected clinicians’ interpretation of the research findings[5]: clinicians were more likely to rate a treatment as beneficial despite the primary outcome being statistically non-significant. Other recent studies have described spin for study designs such as RCTs[6,7], diagnostic accuracy studies[8], non-randomized studies[9] or observational studies[10–12]. These studies have shown that spin used by authors is diverse and highly prevalent.

Authors have highlighted some cases of the distorted interpretation of review findings by authors [13] or misleading reporting that could bias readers’ interpretation of the results[14,15]. Despite the huge impact of systematic reviews and meta-analyses, the interpretation of their findings is still problematic[16–20]. However to our knowledge, we lack a classification scheme for spin for systematic reviews and meta-analyses.

We aimed to 1) develop a classification scheme for spin in reports of systematic reviews and meta-analyses and 2) rank spin that could occur in report abstracts according to their perceived severity (i.e., the likelihood of distorting readers' interpretation).

MethodS

Definition

We defined spin as a specific way of reporting, deliberate or not, to convince readers that the beneficial effect of the experimental treatment in terms of efficacy or safety is greater than that shown by the results (i.e., overstate efficacy and/or understate harm). In this definition we did not consider spin that understated the beneficial effect of the experimental intervention in terms of safety and efficacy (i.e., understate efficacy and/or overstate harm).

We defined the severity of spin as the likelihood of distorting readers' interpretation of the results.

Development of a classification scheme of spin in systematic reviews and meta-analyses

We developed a classification scheme of spin in systematic reviews and meta-analysis of therapeutic interventions in a four-phase consensus process.

A steering group with a high level of expertise in conducting and reporting systematic reviews was responsible for the planning and supervision of this study. The steering group included all the authors of the present study.

Phase 1: We performed a literature review of misleading, biased, misconducted or misinterpreted systematic reviews and meta-analyses to identify ways of reporting that could overstate efficacy or understate harm, which we defined as spin. We searched several sources: 1) the Cochrane Handbook, part 2: "general methods for Cochrane review"[21]; 2) the Methodological standards for the conduct of Cochrane Intervention Review version 2

(MECIR) [22]; 3) MEDLINE and the Cochrane Methodology Register to identify commentaries, letters, editorials or clinical trials; and 4) main articles with recommendations about systematic reviews and meta-analysis interpretation [23–29]. More details are available in the appendix.

Phase 2: Two authors (AY and IB) classified potential spin retrieved in phase 1 according to their location in the article (main text or abstract, including the title) to develop a preliminary classification scheme.

Phase 3: Interviews face to face or by conference call were organized with all members of the steering committee. After a presentation of the concept of spin, we presented the preliminary classification scheme developed in phase 2. Then, in an open discussion, we discussed each potential type of spin, during which clarifications, rewording, opinions, justifications or new ideas were expressed.

Phase 4: The preliminary classification of spin was refined, taking into account the commentary addressed in phase 3. Then we obtained a final classification of spin for systematic reviews and meta-analyses of therapeutic interventions. We re-sent the classification scheme for approval to the steering committee.

Ranking spin in the abstract according to its perceived severity

Study design

We ranked spin in the article abstract (including the title) according to its perceived severity by using the Q-sort methodology. This research method, first described by William

Stephenson in 1935, combines qualitative and quantitative methods and provides a scientific foundation for studying subjectivity or a person's viewpoint[30–33].

The survey focused on spin in article abstracts because this is the most important section of the article. Actually, abstracts are often the only freely available part of the study and are the first and sometimes the only source of information for readers[34,35]. Indeed the dissemination of study findings mainly results from the information provided in the abstract.

Ranking involved 3 steps. Each example of spin was reported on a virtual “card”. First, participants had to read the cards carefully and divide them into 3 piles according to their severity: 1) less severe, 2) medium severity and 3) more severe. Second, participants were asked to take the cards from each pile (less severe / medium severity / more severe), read them again and place them on a score sheet ranging from -3, less severe, to +3, more severe. Once they had placed all cards on the score sheet, they could review the distribution once more and shift the cards one last time. The survey was pilot-tested with a convenience sample to determine whether all cards were understandable. We performed the survey using FlashQ Software 1.0, the Flash application for performing Q-sorting online[36].

All details of the steps, with screenshots of the online survey, are in the appendix.

Participants

We invited the following members of the Cochrane Collaboration to participate in the survey:

1) the coordinating editors in any group, 2) the members of the Methodology Review Group and Bias Method Group, 3) all editors of review groups that published the most reviews

between 2010 and 2013: musculoskeletal, pregnancy and childbirth, anesthesia, neonatal, airway and pain palliative and supportive care groups.

A standardized and personalized e-mail was used to invite all potential participants to participate in a survey to sort spin in article abstracts according to the perceived severity of the spin defined as the likelihood that the spin distorted the interpretation of the results. The email is available in the appendix. We sent 1 or 2 reminders in absence of a response to the previous e-mail.

Analysis

Means and standard deviations were used to describe quantitative variables and frequencies and percentages to describe categorical variables. Spin was classified by mean rank. Statistical analysis involved SAS v9.3 (SAS Inst. Inc., Cary, NC).

RESULTS

Identification of spin in systematic reviews and meta-analyses

From our literature review and consensus process, we identified 39 different types of spin: 28 could occur in the main text and 21 in the abstract. From the 39 types of spin, 13 (33%) were specific to systematic reviews and meta-analyses (Table 1) and 26 (67%) could occur in other study designs. For example, the extrapolation of study findings from a surrogate marker to the global improvement of the disease could occur in systematic reviews as in RCTs, but the inadequate focus on primary study results instead of meta-analysis results is specific to systematic reviews with a meta-analysis design.

Classification scheme of spin in systematic reviews and meta-analyses

We classified the types of spin into 3 main categories: (1) misleading reporting, (2) misleading interpretation and (3) inadequate extrapolation (Table 1).

1) Misleading reporting was defined as incomplete or inadequate reporting of the methods, study analysis, study results or any important information that could be misleading to the reader. The misleading reporting of systematic reviews included 10 items in the main text and 8 in the abstract (Table 1), such as the selective reporting of efficacy outcome favoring the beneficial effect of the experimental intervention or the lack of reporting the number of studies or the number of patients that contribute to the primary outcome meta-analysis in the abstract. For example, in the systematic review by Punjasawadwong published in 2014 [37], the abstract failed to mention that among the 36 trials included in the systematic review, only 8 studies contributed to the meta-analysis of the primary outcome.

2) **Misleading interpretation** was defined as an interpretation of the study results that could be misleading to the reader. This category contained 13 types of spin for the main text and 10 for the abstract. This category included the interpretation by authors of a non-statistically significant result with a wide confidence interval for harm outcomes as being safe. This category also included the inadequate claim of efficacy of an intervention, despite the high risk of publication bias or heterogeneity. For example, in the systematic review by Wilkes and Navickis[38] comparing albumin and crystalloids in seriously ill patients, the authors concluded that “albumin is a safe treatment,” but the pooled relative risk for death was 1.11 (95% CI, 0.95 to 1.28) for all patients receiving albumin, 1.12 (95% CI, 0.85 to 1.46) for surgery or trauma patients and 1.76 (95% CI, 0.91 to 2.78) for patients with burns. The authors claimed the safety of albumin based on non-statistically significant results with a wide confidence interval.

3) **Inadequate extrapolation** was defined as an inappropriate generalization of study results. This category contained 5 types of spin for the main text and 3 for the abstract, including, for example, the extrapolation of the results of a specific outcome analysis such as a biological outcome to the global improvement of the disease.

Ranking of spin in abstracts of systematic reviews and meta-analysis according to their severity

Participants

The survey occurred between February 27 and April 10, 2015. We invited 476 experts to take part in the survey, 169 experts (35%) clicked the link at least once and 123 () participated in

the Q-sort process. We excluded one participant that acknowledged a lack of expertise and particular knowledge about the systematic review and meta-analysis design and their interpretation. Half of the participants were located in Europe; 89 (73%) were affiliated with a review group of the Cochrane Collaboration and 51 (42%) with a methods group. The participants had a high level of expertise in systematic reviews; most (n=107, 90%) had been involved at least once in a systematic review as an author (Table 2).

Q-sort survey

Among the 21 types of spin that could occur in the abstract of the systematic reviews (Table 3), the 7 types rated as the most severe were 1) the conclusion formulates recommendations for clinical practice not supported by the findings, 2) the title claims or suggests a beneficial effect of the experimental intervention not supported by the findings, 3) selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention, 4) the conclusion claims safety based on non-statistically significant results with a wide confidence interval, 5) the conclusion claims the beneficial effect of the experimental treatment despite a high risk of bias in primary studies, 6) selective reporting of or overemphasis on harm outcomes or analysis favoring the safety of the experimental intervention and 7) the conclusion extrapolates the review findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covered a class of several interventions). The distribution of responses for the Q-sort process for each type of spin are available in the appendix.

DISCUSSION

To our knowledge, our study is the first to provide a classification scheme of spin in systematic reviews and meta-analyses. We identified 28 types of spin in the main text and 21 in the abstract that were classified in 3 categories: (1) misleading reporting, (2) misleading interpretation and (3) inadequate extrapolation. The ranking of spin in abstract according to their severity highlighted that the most severe spin were spin in title, spin in the conclusion such as the recommendation not supported by the review findings and selective reporting that favor beneficial effect or safety of the experimental treatment.

Our survey focused on spin in abstracts because the dissemination of study findings mainly results from the information provided in the abstract. Readers often base their initial assessment of a trial on the information reported in an abstract, and in some geographic areas, the abstract may be all that health professionals have easy access to[34]. Actually, abstracts are often the only freely available part of the report and are the first and sometimes the only source of information for readers. Actually, the title and abstract conclusions deliver the simplest take-home message that is easy to memorize.

Previous studies have described and proposed a classification of spin for study designs such as RCTs or diagnostic accuracy or observational studies. These studies showed that spin in published articles was common, particularly in the abstract conclusions [4,7,8,39,40]. Furthermore, spin in published articles is disseminated to the public through press releases and news items[6,41] and leads to a misinterpretation of the study findings[42] by readers. Despite the potential important consequence of spin, no study has described spin in systematic reviews and meta-analyses.

In recent years, the biomedical research community widely focused on transparency and reporting quality. The development of reporting guidelines such as the CONSORT [43] statement and extensions for randomized trials or PRISMA[23,44] for systematic reviews was a major step forward for the improvement of scientific reporting. Implementation of these guidelines by the editors allowed an important improvement of the quality of reporting. These guidelines are evidence-based and provide a minimum set of recommendations to facilitate clarity, completeness, and transparency of reporting. However, these reporting guidelines provide very limited recommendations on how results should be interpreted. For example, systematic reviews often provide many analyses; the selection by the authors of the outcomes and the analyses to report in the abstract and to support the conclusion could bias readers' interpretation of the results. Consequently, authors, reviewers and editors may not be sufficiently aware or trained to identify spin.

Moreover, review findings are difficult to interpret[45–47]. Previous studies showed that generating a conclusion whatever the expertise of the authors is difficult. For example, in the study by Lai et al.[47], participants (hospital practitioners attending an Evidence Based Medicine course and final-year medical students) were asked to choose an appropriate conclusion for the abstracts of 4 Cochrane systematic reviews. Most participants were unable to generate the appropriate conclusion in terms of the direction of the effect and strength of evidence. The interpretation of review findings is even more difficult for a consumer audience; Glenton et al. described the difficulties in generating a plain-language summary understandable by the public[48].

Our study has some limitations. First, the ranking of spin in the report abstract reflects the severity perceived by experts in the field of systematic reviews and meta-analyses; we did not

evaluate the real impact of spin on the interpretation of findings by readers. However, the participants, all editors of the Cochrane Collaboration, were experts. Second, we focused on spin in systematic reviews and meta-analyses of therapeutic interventions. Spin may differ in other systematic reviews such as reviews of diagnostic studies or prognostic studies. Finally, our classification is not exhaustive. Indeed, we may not have identified all examples of spin.

Our classification of spin may have a role in preventing and detecting spin that could occur in systematic reviews and meta-analyses. This classification of spin could raise awareness among authors and help editors and reviewers track spin and therefore reduce its use in distorting the interpretation of research findings. Future research is needed to assess the frequency of spin in systematic reviews and meta-analyses and its impact on readers' interpretation of results.

Acknowledgements

We thank all the participants of the survey: Roland Bachter, Ruth Brassington, Mohammed T Ansari, Evan Mayo-Wilson, Yoon K Loke, Wilco Jacobs, Mariska M.G. Leeflang, Lucy Turner, Jennifer Petkovic, Andrea Tricco, Veronica Pitt, Marialena Trivella, Donna Coffin, Sean Grant, Barbara Mintzes, Jarn Wetterslev, Jan Odgaard-Jensen, Julian Higgins, Emma Sydenham, Roberta Scherer, Sunita Vohra, Jordi Pardo Pardo, Matthew Page, Henrica C.W. de Vet, Willi Sauerbrei, Carlos Tajer, Ian Shemilt, Angela Webster, Sarah Nolan, Nicky Cullum, Marie-Pierre Gagnon, Emma Welsh, Tammy Clifford, Steve McDonald, Gustavo Zanolli, Larissa Shamseer, Jelena Savovic, Jennifer Evans, Ian Shrier, Nicole Skoetz, Hayley Jones, Chris Cates, P. Lina Santaguida, M Willson, Katrina Williams, Philippa Middleton, Elisabeth Hodson, Bernd Richter, Jo Weldon, Norma Terrin, Fulvia Baldassarre, Miranda Cumpston, Patrick Archambault, Joseph Mathew, Renea Johnston, Vivian Welch, Wieland, Deborah Caldwell, Kieran, Jessie McGowan, Steff Lewis, Yali Liu, An-Wen Chan, Nuala Livingstone, Eccleston, Karmela Krleza-Jeric, Mersiha Mahmić-Kaknjo, Rebecca Turner, Nader Shaikh, Jasvinder Singh, Elizabeth Tanjong Ghogomu, Harald Herkner, Nathan Pace, Declan Devane, Rachelle Buchbinder, Barney Reeves, Matthias Briel, Miranda Langendam, Guido Schwarzer, Martin Taphoorn, Jani Ruotsalainen, Stephen Morley, Raymond Daniel, Jaime Miranda, Neil O'Connell, Peter Herbison, M J Van Den Bent, Amy Drahota, Kerry Dwan, Michael Bracken, Mia Schmidt-Hansen, Jim Neilson, Kristin Danko, David Haas, Mathew Zacharias, Susan Chang, Janet Wale, Westeel, Gill Gyte, Khadra Galaal, Ewan McNicol, Michael Bennett, Lara Maxwell, Raj Naik, Lee Hooper, Anne Lethaby, Kristin Carson, Erika Ota, Amanda C de C Williams, William B. Hood, Lelia Duley, Sally Spencer, Jesus Lopez Alcalde.

We thank Laura Smales (BioMedEditing, Toronto, Canada) for English language proofreading.

REFERENCES

- [1] Cook DJ, Mulrow CD, Haynes RB. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Ann Intern Med* 1997;126:376–80. doi:10.7326/0003-4819-126-5-199703010-00006.
- [2] 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:e1000326. doi:10.1371/journal.pmed.1000326.
- [3] Fletcher RH, Black B. “Spin” in scientific writing: scientific mischief and legal jeopardy. *Med Law* 2007;26:511–25.
- [4] Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64. doi:10.1001/jama.2010.651.
- [5] 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;32:4120–6. doi:10.1200/JCO.2014.56.7503.
- [6] Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. *PLoS Med* 2012;9:e1001308. doi:10.1371/journal.pmed.1001308.
- [7] Lockyer S, Hodgson R, Dumville JC, Cullum N. “Spin” in wound care research: the reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 2013;14:371. doi:10.1186/1745-6215-14-371.
- [8] Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeftang MMG. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of “spin”. *Radiology* 2013;267:581–8. doi:10.1148/radiol.12120527.
- [9] Prasad V, Jorgenson J, Ioannidis JPA, Cifu A. Observational studies often make clinical practice recommendations: an empirical evaluation of authors’ attitudes. *J Clin Epidemiol* 2013;66:361–6.e4. doi:10.1016/j.jclinepi.2012.11.005.
- [10] Cofield SS, Corona R V., Allison DB. Use of causal language in observational studies of obesity and nutrition. *Obes Facts* 2010;3:353–6. doi:10.1159/000322940.
- [11] Brown AW, Brown MMB, Allison DB. Belief beyond the evidence: Using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *Am J Clin Nutr* 2013;98:1298–308. doi:10.3945/ajcn.113.064410.

- [12] Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009;302:2345–52. doi:10.1001/jama.2009.1757.
- [13] Ezzo J, Bausell B, Moerman DE, Berman B, Hadhazy V. Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *Int J Technol A H C* 2001-4-457-660000.pdf 2001;17:457–66.
- [14] Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006;333:597–600. doi:10.1136/bmj.333.7568.597.
- [15] Cook D, Guyatt G. Colloid use for fluid resuscitation: evidence and spin. *Ann Intern Med* 2001;135:205–8.
- [16] Weed DL. Meta-Analysis and Causal Inference: A Case Study of Benzene and Non-Hodgkin Lymphoma. *Ann Epidemiol* 2010;20:347–55. doi:10.1016/j.annepidem.2010.02.001.
- [17] Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol* 2000;29:387–90. doi:10.1093/ije/29.3.387.
- [18] Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857–65. doi:10.1016/j.jclinepi.2008.03.004.
- [19] Turner EH, Matthews AM, Linardatos E, Tell R a, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252–60. doi:10.1056/NEJMsa065779.
- [20] Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413–5. doi:10.1136/bmj.a117.
- [21] Higgins JPT GS (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]* 2011. about:blank (accessed December 8, 2011).
- [22] Handbook C, Library TC. Methodological standards for the conduct of Cochrane Intervention Reviews Methodological standards for the conduct of Cochrane Intervention Reviews. *Cochrane Database Syst Rev* 2011.
- [23] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- [24] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101–5.

- [25] Langhorne P. Bias in meta-analysis detected by a simple, graphical test. Prospectively identified trials could be used for comparison with meta-analyses. *BMJ* 1998;316:471.
- [26] Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials* 2009;10:43. doi:10.1186/1745-6215-10-43.
- [27] Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525–37. doi:10.1002/sim.1185.
- [28] Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85. doi:10.1186/1745-6215-11-85.
- [29] Sterne JAC, Sutton AJ, Ioannidis JPA, 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.
- [30] Stephenson W. Correlating persons instead of tests. *J Pers* 1935;4:17–24. doi:10.1111/j.1467-6494.1935.tb02022.x.
- [31] Van Exel J, de Graaf G. Q methodology - A sneak preview -. *Soc Sci* 2005;2:1–30.
- [32] Brown SR. Q-Methodology and Quality Research. *Qual Health Res* 1996:561–7.
- [33] Dennis KE. Q methodology: relevance and application to nursing research. *Adv Nurs Sci* 1986;8:6–17.
- [34] Islamaj Dogan R, Murray GC, Névéal A, Lu Z. Understanding PubMed user search behavior through log analysis. *Database (Oxford)* 2009;2009:bap018. doi:10.1093/database/bap018.
- [35] Saint S, Christakis D a, Saha S, Elmore JG, Welsh DE, Baker P, et al. Journal reading habits of internists. *J Gen Intern Med* 2000;15:881–4.
- [36] Hackert C, Braehler G. FlashQ software [Computer software] 2007: <http://www.hackert.biz/flashq/home>.
- [37] Punjasawadwong Y, Phongchiewboon A, Bunchungmongkol N. Bispectral index for improving anaesthetic delivery and postoperative recovery. *Cochrane Database Syst Rev* 2014;6:CD003843. doi:10.1002/14651858.CD003843.pub3.
- [38] Wilkes MM, Navickis RJ. Patient Survival after Human Albumin Administration: A Meta-Analysis of Randomized, Controlled Trials. *Ann Intern Med* 2001;135:149–64.
- [39] Cummings P, Rivara FP. Spin and boasting in research articles. *Arch Pediatr Adolesc Med* 2012;166:1099–100. doi:10.1001/archpediatrics.2012.1461.

- [40] Hróbjartsson A, Gøtzsche PC, Hróbjartsson A, Gotzsche PC. Powerful spin in the conclusion of Wampold et al.'s re-analysis of placebo versus no-treatment trials despite similar results as in original review. *J Clin Psychol* 2007;63:373–7. doi:10.1002/jclp.20357 [doi].
- [41] Sumner P, Vivian-Griffiths S, Boivin J, Williams a., Venetis C a., Davies a., et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *Bmj* 2014;349:g7015–g7015. doi:10.1136/bmj.g7015.
- [42] 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;32:4120–6. doi:10.1200/JCO.2014.56.7503.
- [43] Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:834–40. doi:10.1016/j.jclinepi.2010.02.005.
- [44] Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS Med* 2013;10:e1001419. doi:10.1371/journal.pmed.1001419.
- [45] Henrotin YE. Interpretation of metaanalyses: pitfalls should be more widely recognized. *J Rheumatol* 2012;39:1107–9. doi:10.3899/jrheum.111420.
- [46] Ioannidis JP a. Meta-research: The art of getting it wrong. *Res Synth Methods* 2010;1:169–84. doi:10.1002/jrsm.19.
- [47] Lai NM, Teng CL, Lee ML. Interpreting systematic reviews: are we ready to make our own conclusions? A cross-sectional study. *BMC Med* 2011;9:30. doi:10.1186/1741-7015-9-30.
- [48] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Med Decis Making* 2010;30:566–77. doi:10.1177/0272989X10375853.

Contributions

Conceived, designed and experiments: AY, IB, DM, DA, TL, AH, PR, Wrote the first draft: AY IB, Contributed to the writing of the manuscript: AY, IB, DM, DA, TL, AH, PR

On-line survey design and data management: Isabelle Pane, senior computer engineer Centre de recherche Épidémiologies et Biostatistiques, INSERM U1153, Paris, France

Data analyses : Gabriel Baron, senior statistician, Centre de recherche Épidémiologies et Biostatistiques, INSERM U1153, Paris, France

Transparency declaration

AY affirms that this 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.

AY and IB had access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of interest

None of the authors have conflicts of interest to declare.

Financial support

AY was funded by a grant from the Fondation pour la Recherche Médicale (FDT20140930875).

Table 1: Classification of spin in the main text and abstracts of systematic reviews and meta-analyses. Spin highlighted in blue are specific to systematic review ad meta-analyses design.

Main category	Description of spin in the main text	Description of spin in abstract
Misleading reporting	<ol style="list-style-type: none"> 1) Failure to acknowledge a departure from protocol that could modify the interpretation of results 2) Selective reporting of or overemphasis on efficacy outcomes favoring the beneficial effect of the experimental intervention (e.g., secondary outcomes, subgroup analyses) 3) Selective reporting of or overemphasis on harm outcomes favoring the safety of the experimental intervention 4) No or inadequate reporting of the limitations of the systematic review 5) Selective citation of articles in favor of the beneficial effect of the experimental intervention 6) Authors hide or do not present any conflict of interest 7) Conclusion focusing selectively on statistically significant efficacy outcome 8) Selective reporting of analysis favoring the beneficial effect of the experimental intervention (e.g., selective analysis using random or fixed effect according the results) 9) Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results 10) Changing the scale of the forest plot to magnify the results (diamond size) 	<ol style="list-style-type: none"> 1) In the abstract, selective reporting of or overemphasis on efficacy outcomes favoring the beneficial effect of the experimental intervention 2) In the abstract, selective reporting of or overemphasis on harm outcomes favoring the safety of the experimental intervention 3) In the abstract, failure to report a wide confidence interval of estimates 4) Authors hide or do not present any conflict of interest 5) Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results 6) Conclusion focusing selectively on statistically significant efficacy outcome 7) In the abstract, failure to report the number of studies/patients actually contributing to the analysis for main outcomes 8) In the abstract, failure to specify the direction of the effect when it favors the control intervention
Misleading interpretation	<ol style="list-style-type: none"> 11) Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings 12) Inadequate interpretation of non-statistically significant results (with a wide confidence interval) as a lack of effect or an equivalent effect for efficacy outcomes 13) Inadequate interpretation of non-statistically significant results (with a wide confidence interval) as 	<ol style="list-style-type: none"> 9) Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings 10) Inadequate focus on <i>p</i> value instead of magnitude of the effect estimates for harm or efficacy outcome 11) Focus on relative effect when the absolute effect is small 12) Conclusion claiming equivalence or comparable effectiveness for non-statistically significant results with a wide

	<p>demonstrating safety for harm outcome</p> <p>14) Inadequate focus on <i>p</i> value instead of magnitude of the effect estimates for harm or efficacy outcome</p> <p>15) Focus on relative effect when the absolute effect is small</p> <p>16) Misleading interpretation of cited articles, favoring the beneficial effect of the experimental intervention</p> <p>17) Conclusion claiming equivalence or comparable effectiveness for non-statistically significant results with a wide confidence interval</p> <p>18) Conclusion formulating recommendations for clinical practice not supported by the findings</p> <p>19) Conclusion claiming safety based on non-statistically significant results with a wide confidence interval</p> <p>20) Conclusion ignoring the high risk of bias of the studies, the heterogeneity or the reporting bias (i.e., a low level of evidence) in the interpretation of the results</p> <p>21) No or inadequate consideration of heterogeneity in results interpretation (i.e., no assessment of heterogeneity reported, claiming the absence of heterogeneity not supported by the data, claiming the beneficial effect of the treatment despite high heterogeneity, no downgrading the evidence in cases of high heterogeneity, interpreting non-statistical significant results for the test of heterogeneity as an evidence of no heterogeneity etc.)</p> <p>22) No or inadequate consideration of the risk of bias of primary studies included in results interpretation (i.e., no reporting of the risk of bias of the primary studies, claiming the low risk of bias of studies included not supported by the data, no downgrading the evidence despite several high risk of bias studies)</p> <p>23) No or inadequate consideration of reporting bias in results interpretation (i.e., no reporting of an assessment of reporting bias, claiming efficacy despite an evidence of reporting bias, claiming the absence of reporting bias</p>	<p>confidence interval</p> <p>13) Conclusion formulating recommendations for clinical practice not supported by the findings</p> <p>14) Conclusion claiming safety based on non-statistically significant results with a wide confidence interval</p> <p>15) Conclusion claiming the beneficial effect of the experimental treatment despite high risk of bias in primary studies</p> <p>16) Conclusion claiming the beneficial effect of the experimental treatment despite reporting bias</p> <p>17) Conclusion claiming the beneficial effect of the experimental treatment despite high heterogeneity</p> <p>18) In the abstract, ignoring that the review included different study design (e.g., controlled trial or observational studies</p>
--	--	---

	not supported by the data, negative test result interpreted as absence of publication bias, use of the test without the condition of validity, inadequate interpretation of a funnel plot, etc.)	
Misleading Extrapolation	<p>24) Inadequate extrapolation of the results from surrogate markers or specific outcome to the global improvement of the disease</p> <p>25) Inadequate extrapolation of the results to a larger population, a larger setting or a wider set of interventions (e.g., from a specific rehabilitation program to all rehabilitation programs, to a specialized unit to a non-specialized medical unit etc.)</p> <p>26) Conclusion extrapolating the review's findings to a different population or setting</p> <p>27) Conclusion extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)</p> <p>28) Conclusion extrapolating the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease</p>	<p>19) Conclusion extrapolating the review's findings to a different population or setting</p> <p>20) Conclusion extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)</p> <p>21) Conclusion extrapolating the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease</p>

Table 2: General characteristics of the expert participants in the survey

Characteristics		Total N=122
Location	Europe	58 (49)
	Canada	24 (20)
	Oceania	15 (13)
	USA	12 (10)
	Asia	7 (6)
	South America	2(2)
Cochrane group member	Primary review group	69 (56)
	Methods group	51 (42)
	Both	20 (16)
	No answer	2 (2)
No. of systematic reviews involved in	0	12 (10)
	1-3	54 (45)
	4-6	97 (31)
	7-10	5 (4)
	>10	11(9)
No. of systematic reviews peer reviewed	0	16 (13)
	1-3	46 (39)
	4-6	29 (24)
	7-10	12 (10)
	>10	16 (13)

Data are no. (%).

Table 3: Spin in abstracts ranked by severity, Q-sort results

Label	Mean rating of severity	Rank
Conclusion contains recommendations for clinical practice not supported by the findings	1.83 (1.18)	1
Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings	1.43 (1.49)	2
Selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention	0.75 (1.20)	3
Conclusion claims safety based on non-statistically significant results with a wide confidence interval	0.46 (1.18)	4
Conclusion claims the beneficial effect of the experimental treatment despite high risk of bias in primary studies	0.43 (1.32)	5
Selective reporting of or overemphasis on harm outcomes or analysis favoring the safety of the experimental intervention	0.31 (1.45)	6
Conclusion extrapolates the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)	0.19 (1.27)	7
Conclusion extrapolates the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease	0.16 (1.42)	8
Conclusion claims the beneficial effect of the experimental treatment despite reporting bias	-0.12 (1.33)	9
Authors hide or do not present any conflict of interest	-0.16 (1.78)	10
Conclusion focuses selectively on statistically significant efficacy outcome	-0.17 (1.54)	11
Conclusion claims equivalence or comparable effectiveness for non-statistically significant results with a wide confidence interval	-0.19 (1.40)	12
Failure to specify the direction of the effect when it favors the control intervention	-0.2 (1.62)	13
Failure to report a wide confidence interval of estimates	-0.29 (1.28)	14
Conclusion extrapolates the review's findings to a different population or setting	-0.43 (1.43)	15
Ignores that the review included different study design (e.g., controlled trial or observational studies)	-0.44 (1.40)	16
Conclusion claims the beneficial effect of the experimental treatment despite high heterogeneity	-0.51 (1.24)	17
Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results	-0.63 (1.55)	18
Ignores the number of studies/patients actually contributing to the analysis for main outcomes	-0.69 (1.39)	19
Focus on <i>p</i> value instead of magnitude of the effect estimates for harm or efficacy outcome	-0.78 (1.20)	20
Focus on relative effect when the absolute effect is small	-0.96 (1.34)	21

Data are mean (SD) rating, ranging from -3, less severe, to +3, more severe.

Figure 1: Example of sorting

LESS SEVERE			MORE SEVERE			
-3	-2	-1	0	+1	+2	+3
(20) Conclusion claiming the beneficial effect of the experimental treatment despite reporting bias.	(1) Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings.	(6) Selective reporting of or overemphasis on harm outcomes or analysis favoring the safety of the experimental intervention.	(18) Conclusion claiming the beneficial effect of the experimental treatment despite high risk of bias in primary studies.	(19) Conclusion claiming the beneficial effect of the experimental treatment despite high heterogeneity.	(2) Ignoring the number of studies / patients actually contributing to the analysis for main outcomes.	(16) Conclusion, extrapolating the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease.
	(9) Failure to report a wide confidence interval of estimates.	(5) Selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention.	(11) Conclusion, focusing selectively on statistically significant efficacy outcome.	(3) Ignoring that the review included different study design (e.g., controlled trial or observational studies).	(21) Authors hide or do not present any conflict of interest.	
	(13) Conclusion, claiming safety based on non-statistically significant results with a wide confidence interval.	(8) Focus on relative effect when the absolute effect is small.	(14) Conclusion, formulating recommendations for clinical practice not supported by the findings.	(15) Conclusion, extrapolating the review's findings to a different population or setting.	(7) Focus on p-value instead of magnitude of the effect estimates for harm or efficacy outcome.	
		(17) Conclusion, extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of...	(4) Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results.	(12) Conclusion, claiming equivalence or comparable effectiveness for non-statistically significant results with a wide confidence interval.		
			(10) Failure to specify the direction of the effect when it favors the control intervention.			

Continue...

Pink, -3 to -1 severity; grey, 0 severity; green, +1 to +3 severity.

5. Discussion et perspectives

5.1. Discussion

5.1.1.Principaux résultats

Les revues systématiques et les méta-analyses sont aujourd'hui un des éléments indispensables de l'évaluation thérapeutique. Elles sont considérées comme apportant le plus haut niveau de preuve[168] ; les recommandations de pratique clinique ainsi que les décisions médicales s'appuient directement sur leurs résultats. Cependant, l'interprétation des données issues des revues systématiques et des méta-analyses est difficile. Les auteurs doivent rapporter de manière transparente et juste tous les éléments permettant aux lecteurs de faire leur propre analyse critique. De ce fait les lecteurs pourront interpréter au mieux les études.

Ce travail de thèse a permis d'étudier 1) la diffusion des *spin* dans les médias et leur impact sur l'interprétation des résultats par les lecteurs, 2) l'impact de la section limite dans les résumés de revues systématiques et de méta-analyses sur l'interprétation par les lecteurs et 3) le développement d'une classification des *spin* des revues systématiques et des méta-analyses et leur classement selon leur probabilité à fausser l'interprétation des lecteurs.

Le premier travail de thèse a permis de montrer que les *spin* étaient fréquents dans les communiqués de presse rapportant des ERC. Nous avons montré que l'existence de *spin* dans les communiqués et articles de presse était fortement associée d'une part à la présence de *spin* dans la publication de l'ERC (RR = 5.6, IC 95% [2.8 - 11.1], p<0.001) et d'autre part à une interprétation biaisée des résultats de l'ERC.

Dans le deuxième travail de thèse nous avons réalisé un essai randomisé contrôlé pour étudier l'impact de l'ajout d'une section limite dans le résumé des revues systématiques sur l'interprétation des lecteurs. Dans notre étude, l'ajout d'une section limite écrite selon les

recommandations « PRISMA for abstract » n'a pas modifié la confiance accordée par les lecteurs dans les résultats de la revue.

Dans le dernier travail de thèse, nous avons développé une classification des *spin* dans les revues systématiques et les méta-analyses. Nous avons ensuite classé les *spin* des résumés selon leur probabilité à fausser l'interprétation des lecteurs.

5.1.2.Place de nos travaux dans la recherche

Mon premier travail de thèse sur la diffusion des *spin* dans les communiqués et articles de presse a permis de sensibiliser la communauté scientifique à cette problématique. En effet, notre premier article de thèse a fait l'objet de plusieurs éditoriaux [169][170], et depuis sa publication en 2012 celui-ci fut cité 26 fois. Deux types de publications ont fait suite à notre étude. Premièrement, des auteurs ont décrit les *spin* retrouvés dans les publications dans leur propre domaine de recherche, comme par exemple les *spin* dans le domaine de la chirurgie plastique[171], de la radiologie[170] ou de l'épidémiologie[172]. Deuxièmement, des chercheurs se sont intéressés aux messages délivrés par les communiqués de presse et à la diffusion de ces messages à travers les médias[173–176]. Un article publié en 2014 dans le *BMJ* a confirmé nos résultats[167]. Dans cette cohorte de 462 communiqués de presse édités par 20 universités anglaises, les principales stratégies de *spin* utilisées par les auteurs et leur fréquence étaient : 1) des conseils ou des recommandations exagérés, 40% (IC 95% [33 - 46]), 2) l'utilisation de termes de causalités inappropriés, 33% (IC 95% [26 - 40]) et 3) l'extrapolation de données animales à l'homme 36% (IC 95% [28 - 46]). La présence de *spin* dans les articles de presse était *fortement* associée à la présence de *spin* dans les communiqués de presse avec des odds ratio compris entre 6,5 et 56 selon le *spin* considéré.

Nos travaux sur les *spin* ont également permis de sensibiliser le grand public et les journalistes. Notre article a été repris dans de nombreux médias : articles de journaux (*the Scientist, the Journalist's Ressource*), blogs et réseaux sociaux (score Altmetric = 412).

Mon deuxième projet de thèse a montré qu'il était possible de réaliser un essai randomisé contrôlé dans le domaine de la méthodologie. Ce type d'étude est rarement utilisé dans notre domaine de recherche, pourtant il s'agit d'un plan d'étude apportant un haut niveau de preuve et permettant d'établir un lien de causalité. I. Boutron a également utilisé un essai randomisé contrôlé pour montrer l'impact des *spin* dans les résumés d'essais randomisés contrôlés dans le domaine de la cancérologie. Pour la première il a ainsi été montré que la présence de *spin* dans les résumés d'articles entraînait une modification de l'interprétation des résultats par les lecteurs[145].

5.1.3.Implications

Une interprétation juste des résultats scientifiques par les lecteurs, professionnels ou non, est indispensable pour assurer la transposition des résultats de la recherche à la pratique clinique et ainsi améliorer l'efficacité et la sécurité des interventions thérapeutiques. En effet une interprétation erronée des résultats scientifiques peut compromettre la prise de décision dans le domaine de la santé des patients. Améliorer la transparence de la recherche et son intégrité est un enjeu majeur de santé publique.

5.1.3.1. Le rôle de la communauté scientifique dans la diffusion des spin

Nos travaux ont montré que les *spin* étaient fréquents dans les RCT et qu'ils étaient associés à la présence de *spin* dans les communiqués et articles de presse. Ce résultat est très important car il a permis de remettre en cause une idée fréquemment répandue, que les journalistes seraient à l'origine des messages erronés dans les articles de vulgarisation scientifiques. En effet, plusieurs articles ont montré que les communiqués de presse et les articles de presse destinés au grand public portaient un message erroné, parce que les limites des études n'étaient pas décrites, les résultats insuffisamment détaillés ou bien largement extrapolés[50,177,178]. Les résultats de notre étude montrent que les *spin* présents dans les articles de presse n'ont pas été « rajoutés » par les journalistes, mais était déjà présents dans la publication scientifique ou le communiqué de presse. Ainsi l'interprétation erronée des résultats scientifiques diffusée par les médias provient au moins en partie des auteurs des articles scientifiques.

L'écriture d'un manuscrit rapportant des résultats scientifiques constitue l'aboutissement d'un projet de recherche. L'utilisation de *spin* dans l'écriture scientifique peut résulter de l'ignorance des standards scientifiques, du mimétisme des plus jeunes envers leur aînés, de préjugés inconscients ou de l'intention délibérée d'influencer le lecteur[144]. Ce type de comportement peut être par ailleurs favorisé par la forte pression à laquelle sont soumis les chercheurs pour être publiés (« publish or perish »)[179].

Nos résultats mettent ainsi en avant la responsabilité de la communauté scientifique dans la diffusion des résultats de la recherche et des messages transmis au public.

5.1.3.2. Améliorer la communication des résultats scientifiques: un besoin de nouveaux outils

L'écriture d'un article scientifique n'est pas simplement un rapport des faits. Les lecteurs ne sont pas toujours des spécialistes de la question scientifique ou de la méthodologie utilisée ; ils ont besoin d'informations pour comprendre le contexte et les enjeux que les résultats de l'étude impliquent. Aussi les rôles d'un article scientifique sont: 1) de décrire précisément et clairement la méthode employée afin de permettre la réplication de l'étude par une équipe tierce, 2) de rapporter tous les résultats de l'étude en termes d'efficacité et de sécurité, et 3) de donner aux lecteurs des informations sur le contexte de la recherche, comme par exemple les autres études réalisées dans le même domaine de recherche, les limites de l'étude ou des explications permettant de comprendre les implications de tel ou tel résultat.

La qualité des publications s'est améliorée depuis l'essor des *reporting guidelines*. L'étude de S. Hopwell[180] a comparé la description des caractéristiques méthodologiques dans les

publications d'ERC entre 2000 et 2006, soit avant et après l'implémentation des recommandations CONSORT en 2001. Cette étude a montré une amélioration de la description de certains points méthodologiques comme la description du critère principal de jugement, le calcul du nombre de sujets nécessaires ou la méthode utilisée pour obtenir la séquence de randomisation, (figure 4).

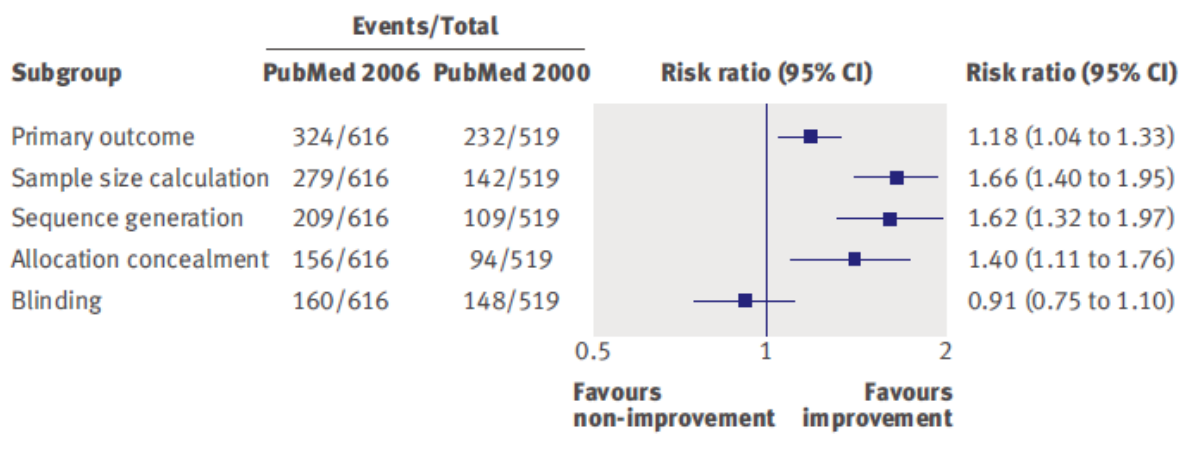


Figure 4: Les différences de description des éléments méthodologiques entre 2000 et 2006 [180]

Ces initiatives ont permis d'améliorer la qualité du *reporting*, cependant elles ne prennent pas en compte la présentation et l'interprétation des résultats par les auteurs. En effet, des études qui suivent scrupuleusement les recommandations peuvent tout de même contenir des *spin*. Par exemple dans un ERC comparant deux seuils transfusionnels, restrictif ou libéral, en chirurgie cardiaque, publié dans le *New England Journal of Medicine* en Mars 2015, le critère principal de jugement- la morbidité- était statistiquement non significatif (OR = 1,11 ; IC 95% [0,91-1,34], p=0,3). La mortalité, critère de sécurité pré-spécifié comme étant un critère secondaire, était statistiquement significative avec un sur-risque de mortalité dans le groupe expérimental - seuil transfusionnel restrictif - (OR = 1,64 ; IC 95% [1,00-2,67], p=0,045). La

conclusion du résumé prend en compte exclusivement le résultat du critère principal de jugement, la morbidité, sans tenir compte du critère de mortalité. Les auteurs concluent : « a restrictive transfusion threshold after cardiac surgery was not superior to a liberal threshold with respect to morbidity or health care costs ». Cette conclusion est, à mon sens, trompeuse car elle ne permet pas de mettre en évidence un résultat majeur de sécurité.

De même, dans la revue systématique avec méta-analyses de Y. Punjasawadwong, publiée par la *Cochrane Database of Systematic Reviews* en 2014, les auteurs rapportent dans le résumé le nombre d'essais inclus dans la revue systématique, 36 essais, mais ne mentionnent pas le nombre d'essais ayant contribué aux méta-analyses des critères répondant à l'objectif principal, soit huit essais [181]. Cette revue systématique évaluait le bénéfice de l'utilisation per opératoire du BIS sur le risque de réveil préopératoire, la consommation d'agents anesthésiques, et le délai de réveil. Dans cette étude, les auteurs mettent en évidence un potentiel biais de publication et une hétérogénéité entre les études, seule une petite proportion des études incluses participant à l'analyse principale et la qualité de la preuve est décrite comme modérée à faible. Pourtant dans le résumé destiné au grand public les auteurs affirment l'utilité du BIS : « The results from this updated review indicate that BIS can be useful in guiding the anaesthetic dose to avoid the risk of intraoperative awareness in surgical patients at high risk for awareness. Furthermore, anaesthesia guided by BIS improves anaesthetic delivery and recovery from anaesthesia ».

Dans le deuxième travail de cette thèse, nous avons montré que l'ajout d'informations supplémentaires portant sur les limites de la revue n'a pas eu d'incidence sur l'interprétation globale des résultats par des lecteurs-experts. La section limite créée était rapportée de façon factuelle et neutre, et nous n'avons pas modifié la conclusion du résumé. Aussi l'interprétation des résultats par les lecteurs est un phénomène complexe qui dépend à la fois des éléments

rapportés dans l'article, mais aussi de la manière dont ils sont présentés et interprétés par les auteurs. Les *reporting guidelines* ne précisent pas comment doivent être présentés et surtout interprétés les résultats des études. Par exemple, l'item 26 des recommandations PRISMA concernant la conclusion stipule : « La conclusion doit fournir une interprétation générale des résultats dans le contexte des preuves existantes et doit donner les implications pour la recherche future ». Mais en aucun cas les recommandations ne précisent comment et sur quels critères la conclusion doit s'appuyer. Ainsi, l'application des *reporting guidelines* actuelles ne permettent pas d'éviter les *spin*.

Malgré leur fréquence élevée, et leur diffusion importante la reconnaissance des *spin* dans les publications est difficile même pour un public averti. Les classifications de *spin* développées dans les différents domaines de recherche ou pour différents types d'étude permettent de sensibiliser les lecteurs et peuvent ainsi servir d'outil. La connaissance des *spin* potentiellement présents dans les publications permet d'une part d'être plus vigilant lors de la lecture d'un article et d'autre part de ne pas être influencé en leur présence. C'est pourquoi il est important de diffuser ces classifications et de former les médecins tant lors de leur formation initiale qu'en formation continue.

Comme nous l'avons vu précédemment, l'interprétation des résultats des revues systématiques est difficile même pour des lecteurs expérimentés et/ou ayant suivis une formation à l'EBM. Les experts qui participent à l'élaboration des recommandations de pratique clinique sont également potentiellement influencés par la présence de *spin* dans les publications. Il est donc important que les sociétés savantes s'emparent également de ce concept nouveau et organisent des formations spécifiques pour les praticiens et en particulier pour ceux qui participent aux recommandations de pratique clinique.

Dans le troisième travail de cette thèse nous avons développé une classification des *spin* dans les revues systématiques et les méta-analyses, et nous avons identifié les *spin* qui semblaient avoir la plus grande probabilité de fausser l'interprétation des lecteurs.

En sensibilisant les auteurs, les éditeurs et les relecteurs, notre classification pourrait avoir un rôle important dans la prévention et la détection des *spin* dans les revues systématiques et les méta-analyses. La promotion des check-lists comme celle du CONSORT et du PRISMA sont des aides importantes pour l'amélioration de la qualité des articles[180]. La promotion de notre classification des *spin* dans les revues systématiques pourrait également être un support pour les auteurs en les informant sur les mauvaises pratiques à éviter et en les sensibilisant au mode de présentation et d'interprétation des études.

Aussi les éditeurs et les relecteurs doivent avoir un rôle de garde-fou. Le classement des *spin* selon leur sévérité permet de hiérarchiser les *spin* les plus sévères et donc ceux qui doivent être évités à tout prix.

L'élaboration de la classification des *spin* dans les revues systématiques et les méta-analyses et leur classement selon leur sévérité sont une première étape. De futurs travaux de recherche sont nécessaires pour déterminer la fréquence des *spin* dans les revues systématiques et les méta-analyses ainsi que leur impact sur l'interprétation des résultats par les lecteurs.

5.2. Conclusion

Les *spin* sont un obstacle à la transposition des résultats de la recherche à la pratique clinique courante ; ils sont fréquents dans les publications scientifiques et leurs conséquences potentielles sont importantes. Les *spin* sont largement diffusés au-delà des revues

scientifiques, ils touchent ainsi un large public et peuvent être à l'origine d'une interprétation erronée des résultats de la recherche avec de potentielles implications sur la pratique clinique.

Les revues systématiques sont devenues très populaires et très utiles. Les cliniciens, les agences d'évaluation de santé et les organismes de régulation utilisent de plus en plus ce mode de synthèse. Cependant, comme nous l'avons mis en évidence, leur interprétation est difficile et la présence de *spin* peut potentiellement biaiser l'interprétation des résultats par les lecteurs.

5.3. Perspectives

Nous allons poursuivre nos travaux de recherche selon plusieurs axes.

5.3.1. Fréquence, diffusion et impact des *spin* dans les revues systématiques et les méta-analyses

Nous allons poursuivre notre travail de description et d'investigation des *spin* dans les revues systématiques et les méta-analyses. Nous allons évaluer la fréquence des *spin* dans les revues systématiques et les méta-analyses, leur impact et leur diffusion. Pour cela nous allons sélectionner un échantillon représentatif de revues systématiques et de méta-analyses publiées récemment. Nous allons chercher de manière systématique la présence de *spin* dans les revues sélectionnées. Cela nous permettra également d'enrichir notre classification des *spin*. Ensuite, nous évaluerons l'impact des *spin* dans le résumé des revues systématiques sur l'interprétation des lecteurs à l'aide d'un essai randomisé contrôlé. Et enfin, nous étudierons la diffusion des *spin* présents dans les ERC et les revues systématiques et leurs impacts sur la pratique clinique. Pour cela nous souhaitons évaluer l'impact des *spin* lors de l'élaboration des recommandations de pratique clinique par des experts cliniciens.

5.3.2. Développement d'outils d'aide à la prévention et à la détection des *spin*

Nous souhaitons développer un outil d'aide à l'écriture pour les auteurs de revues systématiques : « Writing Aid Tool ». Pour développer cet outil, nous utiliserons les recommandations et classifications des *spin* dans les revues systématiques et les méta-analyses précédemment développées. Cet outil pourrait être implémenté dans le logiciel RevMan, avec pour objectif de prévenir les *spin* au moment de la rédaction de la revue par les auteurs. Cet outil consistera en l'apparition de fenêtres pop-up au moment de la rédaction de la revue dans le logiciel RevMan. Chaque fenêtre pop-up contiendrait des recommandations pour l'interprétation des résultats ainsi que les principaux *spin* à éviter. Par exemple, lors de la rédaction de la section conclusion du résumé, l'auteur de la revue aurait juste à cliquer sur une icône pour voir une fenêtre s'ouvrir décrivant les *spin* spécifiques à cette section.

Nous souhaitons également développer un outil pour identifier les *spin* dans les revues systématiques et les méta-analyses, destiné aux relecteurs et aux éditeurs ainsi que des outils pédagogiques destinés aux étudiants lors de leur formation initiale mais également aux médecins lors de leur formation continue.

5.3.3. De nouveaux outils pour améliorer la communication des résultats des revues systématiques

Des améliorations quant au contenu ou à la structure des résumés ont déjà été proposées principalement dans le cadre des ERC. Le *JAMA* a innové en 2013 en publiant pour la première fois des résumés avec des tableaux de résultats[183,184]. De même, la maison d'édition Elsevier propose aux auteurs de résumer les principaux résultats d'une étude à l'aide d'une figure, le « graphical abstract ». De tels résumés sont édités dans les revues *International Journal of Pharmaceutics*, le *Journal of Ethnopharmacology* ou dans le *Journal of Molecular Biology*. Cela permettrait potentiellement de diminuer les biais de présentation des résultats puisque ceux-ci seraient présentés de manière plus neutre. Ces nouveaux formats de résumé ne sont pas encore utilisés en routine et posent de nouveaux problèmes comme par exemple l'indexation des tableaux ou figures dans les bases de données comme MEDLINE. Aussi, aucune étude n'a à notre connaissance évalué l'impact de telles modifications de la présentation des résumés d'article sur l'interprétation par les lecteurs.

De même, de nouvelles formes de présentation des résultats dans les résumés des revues systématiques et méta-analyses ont été proposées. Des études émanant de la collaboration Cochrane ont évalué différents modes de présentation des résultats dans le résumé scientifique et dans le résumé adressé au grand public appelé le « plain language summary ». Des modifications du contenu et du format du « plain language summary » ont permis d'améliorer la compréhension par un public non médecin des résultats issus d'une revue systématique avec méta-analyse[182].

Il est important d'étudier l'impact des différentes formes de présentation des résumés sur l'interprétation des résultats par les lecteurs afin de déterminer quel modèle est le plus

efficace. Des recherches méthodologiques sont nécessaires pour développer des outils d'aide à l'écriture des résumés et principalement de sa conclusion. En effet, la partie probablement la plus importante et la plus lue d'un article scientifique est la conclusion du résumé : elle contient le message à retenir, et c'est également elle qui contient le plus de *spin*.

Il n'existe pas à notre connaissance de recommandations qui stipulent comment écrire la conclusion d'une revue systématique. Aussi nous souhaitons élaborer des recommandations pour une écriture standardisée de la conclusion du résumé des revues systématiques. Dans un deuxième temps nous évaluerons l'impact de ce nouveau mode de présentation sur l'interprétation de la revue par les lecteurs.

Bibliographie

- [1] The periodic health examination Canadian Task Force on the Periodic Health Examination. *C Can Med Assoc J* 1979;121:1193–254.
- [2] Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1989;95:2S – 4S.
- [3] Greenhalgh T. *How to Read a Paper: The Basics of Evidence-Based Medicine*. vol. 2010. 2010.
- [4] 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:e1000326. doi:10.1371/journal.pmed.1000326.
- [5] Pearson K. Report on Certain Enteric Fever Inoculation Statistics. *Br J Surg* 1904:1243–6.
<http://www.jstor.org/discover/10.2307/20282622?uid=363394081&uid=3738016&uid=2134&uid=386250441&uid=363394071&uid=2&uid=386250431&uid=70&uid=3&uid=67&uid=62&uid=60&sid=21104837176121> (accessed December 12, 2014).
- [6] 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;268:240–8.
- [7] Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research 2004;328:39–41.
- [8] Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;309:597–9.
- [9] Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA* 1998;280:280–2.
- [10] Roloff V, Higgins JPT, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat Med* 2013;32:11–24. doi:10.1002/sim.5524.
- [11] 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. doi:10.1016/S0140-6736(13)62297-7.
- [12] 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. doi:10.1016/S0140-6736(13)62229-1.

- [13] Ioannidis JPA, 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. doi:10.1016/S0140-6736(13)62227-8.
- [14] Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet* 2010;376:20–1. doi:10.1016/S0140-6736(10)61045-8.
- [15] Cochrane AL. Archie Cochrane in his own words. Selections arranged from his 1972 introduction to “Effectiveness and Efficiency: Random Reflections on the Health Services” 1972. *Control Clin Trials* 1989;10:428–33.
- [16] Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. *Med. year 2000 London Off. Heal. Econ.*, 1979, p. 1–11.
- [17] Higgins JPT GS (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]* 2011. about:blank (accessed December 8, 2011).
- [18] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101–5.
- [19] Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ* 1997;315:1371–4. doi:10.1136/bmj.315.7119.1371.
- [20] Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549. doi:10.1136/bmj.d549.
- [21] Handbook C, Library TC. Methodological standards for the conduct of Cochrane Intervention Reviews Methodological standards for the conduct of Cochrane Intervention Reviews. *Cochrane Database Syst Rev* 2011.
- [22] Kurata K, Morioka T, Yokoi K, Matsubayashi M. Remarkable Growth of Open Access in the Biomedical Field: Analysis of PubMed Articles from 2006 to 2010. *PLoS One* 2013;8:2006–11. doi:10.1371/journal.pone.0060925.
- [23] Barry HC, Ebell MH, Shaughnessy a F, Slawson DC, Nietzke F. Family physicians’ use of medical abstracts to guide decision making: style or substance? *J Am Board Fam Pract* 1998;14:437–42.
- [24] Saint S, Christakis D a, Saha S, Elmore JG, Welsh DE, Baker P, et al. Journal reading habits of internists. *J Gen Intern Med* 2000;15:881–4.
- [25] Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS Med* 2013;10:e1001419. doi:10.1371/journal.pmed.1001419.
- [26] Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts:

- explanation and elaboration. *PLoS Med* 2008;5:e20.
doi:10.1371/journal.pmed.0050020.
- [27] Soffer A. Abstracts of clinical investigations. A new and standardized format. *Chest* 1987;92:389–90.
- [28] Huth E J. Structured Abstracts for Papers Reporting Clinical Trials. *Ann Intern Med* 1987;106:626. doi:10.7326/0003-4819-106-4-626.
- [29] 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. doi:10.1136/bmj.e4178.
- [30] Hopewell S, Clarke M, Askie L. Reporting of trials presented in conference abstracts needs to be improved. *J Clin Epidemiol* 2006;59:681–4.
doi:10.1016/j.jclinepi.2005.09.016.
- [31] Beller EM, Glasziou PP, Hopewell S, Altman DG. Reporting of effect direction and size in abstracts of systematic reviews. *JAMA* 2011;306:1981–2.
doi:10.1001/jama.2011.1620.
- [32] Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437–43.
- [33] Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ* 2007;335:1202–5.
doi:10.1136/bmj.39376.447211.BE.
- [34] Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64. doi:10.1001/jama.2010.651.
- [35] Hopewell S, McDonald S. Grey literature in meta-analyses of randomized trials of health care interventions. *Database Syst Rev* 2007:MR000010.
doi:10.1002/14651858.MR000010.pub3.
- [36] Hopewell S, Boutron I, Altman DG, Ravaud P. Deficiencies in the publication and reporting of the results of systematic reviews presented at scientific medical conferences. *J Clin Epidemiol* 2015;68. doi:10.1016/j.jclinepi.2015.03.006.
- [37] Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL. Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 1991;325:1180–3.
- [38] Schwartz LM, Woloshin S. The media matter: a call for straightforward medical reporting. *Ann Intern Med* 2004;140:226–8. doi:140/3/226 [pii].

- [39] Dentzer S. Communicating Medical News — Pitfalls of Health Care Journalism. *N Engl J Med* 2009;360:1–3. doi:10.1056/NEJMp1002530.
- [40] Jaques H. Get your research reported well in the news. *BMJ* 2011. <http://careers.bmj.com/careers/advice/view-article.html?id=20001805> (accessed April 16, 2015).
- [41] Slaughter AJ, Simon SD. Influence of the news media on diagnostic testing in the emergency department. *Arch Pediatr Adolesc Med* 2003;157:257–60.
- [42] CIHR CIOHR. Of Mice and Men: Making Sense of Health Science in the Media | Canadian Cochrane Centre 2014. <http://ccc.cochrane.org/caf%0c3%a9-scientifique-mice-and-men-making-sense-health-science-media> (accessed April 16, 2015).
- [43] Woloshin S, Schwartz LM. Press releases: translating research into news. *JAMA* 2002;287:2856–8. doi:jbr10380 [pii].
- [44] Van Trigt AM, Haaijer-Ruskamp FM, Willems J, de J den BLT, TFJ T, van Trigt AM, de Jong-van den Berg LT, Haaijer-Ruskamp FM, Willems J, et al. Journalists and their sources of ideas and information on medicines. *Soc Sci Med* 1994;38:637.
- [45] Kuriya B, Schneid EC, Bell CM. Quality of pharmaceutical industry press releases based on original research. *PLoS One* 2008;3:e2828. doi:10.1371/journal.pone.0002828 [doi].
- [46] Entwistle V. Reporting research in medical journals and newspapers. *BMJ* 1995;310:920–3.
- [47] Woloshin S, Schwartz LM, Kramer BS. Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 2009;101:1596–9. doi:djp409 [pii] 10.1093/jnci/djp409 [doi].
- [48] Woloshin S, Schwartz LM, Casella SL, Kennedy AT, Larson RJ. Press releases by academic medical centers: not so academic? *Ann Intern Med* 2009;150:613–8. doi:150/9/613 [pii].
- [49] Woloshin S, Schwartz LM, Casella SL, Kennedy AT, Larson RJ. Press Releases by Academic Medical Centers : Not So Academic ? *Ann Intern Med* 2009;150:613–8.
- [50] Schwartz LM, Woloshin S, Andrews A, Stukel TA. Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study. *BMJ* 2012;344:d8164.
- [51] Shrier I, Boivin J-F, Platt RW, Steele RJ, Brophy JM, Carnevale F, et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* 2008;8:19. doi:10.1186/1472-6947-8-19.

- [52] Lai NM, Teng CL, Lee ML. Interpreting systematic reviews: are we ready to make our own conclusions? A cross-sectional study. *BMC Med* 2011;9:30. doi:10.1186/1741-7015-9-30.
- [53] Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;156:1411–6.
- [54] Fischhoff B, Bostrom A, Quadrel MJ. Risk perception and communication. *Annu Rev Public Health* 1993;14:183–203. doi:10.1146/annurev.pu.14.050193.001151.
- [55] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124–31. doi:10.1126/science.185.4157.1124.
- [56] Lloyd A, Hayes P, Bell PR, Naylor AR. The role of risk and benefit perception in informed consent for surgery. *Med Decis Making* 21:141–9.
- [57] Koehler JJ. The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality. *Organ Behav Hum Decis Process* 1993;56:28–55.
- [58] Panagiotou OA, Ioannidis JPA. Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists. *J Clin Epidemiol* 2012;65:740–7. doi:10.1016/j.jclinepi.2012.01.008.
- [59] Berner ES, Graber ML. Overconfidence as a Cause of Diagnostic Error in Medicine. *Am J Med* 2008;121. doi:10.1016/j.amjmed.2008.01.001.
- [60] Stiegler MP, Tung A. Cognitive Processes in Anesthesiology Decision Making. *Anesthesiology* 2014;120:402–17.
- [61] Chaudhry S, Schroter S, Smith R, Morris J. Does declaration of competing interests affect readers' perceptions? A randomised trial. *BMJ* 2002;325:1391–2.
- [62] Schroter S, Morris J, Chaudhry S, Smith R, Barratt H. Does the type of competing interest statement affect readers' perceptions of the credibility of research? Randomised trial. *BMJ* 2004;328:742–3. doi:10.1136/bmj.38035.705185.F6.
- [63] Kesselheim AS, Robertson CT, Myers J a., Rose SL, Gillet V, Ross KM, et al. A Randomized Study of How Physicians Interpret Research Funding Disclosures. *N Engl J Med* 2012;367:1119–27. doi:10.1056/NEJMsa1202397.
- [64] Silverman GK, Loewenstein GF, Anderson BL, Ubel PA, Zinberg S, Schulkin J. Failure to discount for conflict of interest when evaluating medical literature: a randomised trial of physicians. *J Med Ethics* 2010;36:265–70. doi:10.1136/jme.2009.034496.
- [65] Buffel du Vaure C, Boutron I, Perrodeau E, Ravaud P. Reporting funding source or conflict of interest in abstracts of randomized controlled trials, no evidence of a large

- impact on general practitioners' confidence in conclusions, a three-arm randomized controlled trial. *BMC Med* 2014;12:69. doi:10.1186/1741-7015-12-69.
- [66] Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916–21.
- [67] Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ Br Med J* 2003;327:741–4. doi:10.1136/bmj.327.7417.741.
- [68] Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ* 1995;311:1056–9; discussion 1059–60.
- [69] Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994;343:1209–11.
- [70] Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992;92:121–4.
- [71] Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994;309:761–4.
- [72] Raina PS, Brehaut JC, Platt RW, Klassen TP, Moher D, St John P, et al. The influence of display and statistical factors on the interpretation of metaanalysis results by physicians. *Med Care* 2005;43:1242–9.
- [73] Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24. doi:10.1186/1741-7015-8-24.
- [74] Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest* 2010;40:35–53. doi:10.1111/j.1365-2362.2009.02234.x.
- [75] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- [76] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1–37. doi:10.1016/j.jclinepi.2010.03.004.
- [77] Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Rev Esp Salud Publica* 2007;82:251–9.

- [78] Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;371:281–3. doi:10.1016/S0140-6736(07)61835-2.
- [79] Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol* 2007;60:324–9.
- [80] 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. doi:10.1186/2046-4053-1-60.
- [81] Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M, et al. The reporting quality of meta-analyses improves: a random sampling study. *J Clin Epidemiol* 2008;61:770–5. doi:10.1016/j.jclinepi.2007.10.008.
- [82] De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med* 2004;351:1250–1. doi:10.1056/NEJMe048225.
- [83] DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004;292:1363–4. doi:10.1001/jama.292.11.1363 [doi] 292.11.1363 [pii].
- [84] Bian Z-X, Wu T-X. Legislation for trial registration and data transparency. *Trials* 2010;11:64. doi:10.1186/1745-6215-11-64.
- [85] Zarin D a., Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov Results Database — Update and Key Issues. *N Engl J Med* 2011;364:852–60. doi:10.1056/NEJMsa1012065.
- [86] Booth A, Clarke M, Ghera D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. *Lancet* 2011;377:108–9. doi:10.1016/S0140-6736(10)60903-8.
- [87] Booth A, Clarke M, Ghera D, Moher D, Petticrew M, Stewart L. Establishing a minimum dataset for prospective registration of systematic reviews: an international consultation. *PLoS One* 2011;6:e27319. doi:10.1371/journal.pone.0027319.
- [88] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton a. J, et al. Dissemination and publication of research findings: An updated review of related biases. *Health Technol Assess (Rockv)* 2010;14:1–220. doi:10.3310/hta14080.
- [89] 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:MR000006. doi:10.1002/14651858.MR000006.pub3.

- [90] Kicinski M. Publication bias in recent meta-analyses. *PLoS One* 2013;8:e81823. doi:10.1371/journal.pone.0081823.
- [91] Von Elm E, Röllin A, Blümle 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;138:197–203. doi:2008/13/smw-12027.
- [92] Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640–5.
- [93] Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281–6.
- [94] Tramèr MR, Reynolds DJ, Moore R a, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635–40. doi:10.1136/bmj.315.7109.635.
- [95] Gøtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;36:429–32.
- [96] Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *Thebmj* 2009;339:24. doi:10.1136/bmj.b2680.
- [97] Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987;295:654–6. doi:10.1136/bmj.295.6605.1066-b.
- [98] Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;305:15–9. doi:10.1136/bmj.305.6855.717.
- [99] 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;358:252–60. doi:358/3/252 [pii] 10.1056/NEJMs065779 [doi].
- [100] Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007:MR000005. doi:10.1002/14651858.MR000005.pub3.
- [101] Nguyen T-A-H, Dechartres A, Belgherbi S, Ravaud P. Public availability of results of trials assessing cancer drugs in the United States. *J Clin Oncol* 2013;31:2998–3003. doi:10.1200/JCO.2012.46.9577.
- [102] 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. doi:10.1136/bmj.c365.
- [103] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977–84. doi:302/9/977 [pii] 10.1001/jama.2009.1242 [doi].

- [104] Dickersin K. Reporting and other biases in studies of Neurontin for migrian, psychiatric/bipolar disorders, nociceptive pain, and neuropathic pain n.d. . <http://dida.library.ucsf.edu/pdf/oxx18r10>.
- [105] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials. *JAMA* 2009;302:977–84.
- [106] 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;291:2457–65. doi:10.1001/jama.291.20.2457 [doi] 291/20/2457 [pii].
- [107] Killeen S, Sourallous P, Hunter IA, Hartley JE, Grady HLO. Registration rates, adequacy of registration, and a comparison of registered and published primary outcomes in randomized controlled trials published in surgery journals. *Ann Surg* 2014;259:193–6. doi:10.1097/SLA.0b013e318299d00b.
- [108] Smith SM, Wang AT, Pereira A, Chang RD, McKeown A, Greene K, et al. Discrepancies between registered and published primary outcome specifications in analgesic trials: ACTION systematic review and recommendations. *Pain* 2013;154:2769–74. doi:10.1016/j.pain.2013.08.011.
- [109] 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:MR000031. doi:10.1002/14651858.MR000031.pub2.
- [110] Anand V, Scales DC, Parshuram CS, Kavanagh BP. Registration and design alterations of clinical trials in critical care: a cross-sectional observational study. *Intensive Care Med* 2014;40:700–22. doi:10.1007/s00134-014-3250-7.
- [111] Huić M, Marušić M, Marušić A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. *PLoS One* 2011;6:e25258. doi:10.1371/journal.pone.0025258.
- [112] Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *Br Med J* 2008;337:a2299.
- [113] 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;343:1520–8, 2 p following 1528. doi:10.1056/NEJM200011233432103.
- [114] Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? *BMJ* 2007;334:120–3. doi:334/7585/120 [pii] 10.1136/bmj.39024.487720.68 [doi].

- [115] Hannink G, Gooszen HG, Rovers MM. Comparison of registered and published primary outcomes in randomized clinical trials of surgical interventions. *Ann Surg* 2013;257:818–23. doi:10.1097/SLA.0b013e3182864fa3.
- [116] 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:MR000031. doi:10.1002/14651858.MR000031.pub2.
- [117] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977–84. doi:302/9/977 [pii] 10.1001/jama.2009.1242 [doi].
- [118] Horton R. The rhetoric of research. *BMJ* 1995;310:985–7.
- [119] Latour B, Fabbri P. La rhétorique du discours scientifique : pouvoir et devoir dans un article de sciences exactes. *Actes Rech Sci Soc* 1977;13:81–95.
- [120] Mazzaferro V, Romito R, Schiavo M, Mariani L, Camerini T, Bhoori S, et al. Prevention of hepatocellular carcinoma recurrence with alpha-interferon after liver resection in HCV cirrhosis. *Hepatology* 2006;44:1543–54. doi:10.1002/hep.21415.
- [121] Lockyer S, Hodgson R, Dumville JC, Cullum N. “Spin” in wound care research: the reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 2013;14:371. doi:10.1186/1745-6215-14-371.
- [122] De las Fuentes L, Waggoner AD, Mohammed BS, Stein RI, Miller B V., Foster GD, et al. Effect of Moderate Diet-Induced Weight Loss and Weight Regain on Cardiovascular Structure and Function. *J Am Coll Cardiol* 2009;54:2376–81. doi:10.1016/j.jacc.2009.07.054.
- [123] Nienhuis W a., Stienstra Y, Thompson W a., Awuah PC, Abass KM, Tuah W, et al. Antimicrobial treatment for early, limited Mycobacterium ulcerans infection: a randomised controlled trial. *Lancet* 2010;375:664–72. doi:10.1016/S0140-6736(09)61962-0.
- [124] Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeftang MMG. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of “spin”. *Radiology* 2013;267:581–8. doi:10.1148/radiol.12120527.
- [125] Xu F, Yan Q, Wang H, Niu J, Li L, Zhu F, et al. Performance of detecting IgM antibodies against enterovirus 71 for early diagnosis. *PLoS One* 2010;5:2008–11. doi:10.1371/journal.pone.0011388.
- [126] Cofield SS, Corona R V., Allison DB. Use of causal language in observational studies of obesity and nutrition. *Obes Facts* 2010;3:353–6. doi:10.1159/000322940.

- [127] Brown AW, Brown MMB, Allison DB. Belief beyond the evidence: Using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *Am J Clin Nutr* 2013;98:1298–308. doi:10.3945/ajcn.113.064410.
- [128] Menachemi N, Tajeu G, Sen B, Ferdinand AO, Singleton C, Utley J, et al. Overstatement of results in the nutrition and obesity peer-reviewed literature. *Am J Prev Med* 2013;45:615–21. doi:10.1016/j.amepre.2013.06.019.
- [129] Kenter R, Warmerdam L, Brouwer-Dudokdewit C, Cuijpers P, van Straten A. Guided online treatment in routine mental health care: an observational study on uptake, drop-out and effects. *BMC Psychiatry* 2013;13:43. doi:10.1186/1471-244X-13-43.
- [130] Cummings P, Rivara FP. Spin and boasting in research articles. *Arch Pediatr Adolesc Med* 2012;166:1099–100. doi:10.1001/archpediatrics.2012.1461.
- [131] Wilson JR. Rhetorical strategies used in the reporting of implantable defibrillator primary prevention trials. *Am J Cardiol* 2011;107:1806–11. doi:10.1016/j.amjcard.2011.02.320.
- [132] Bartlett C, Sterne J, Egger M. What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *BMJ* 2002;325:81–4.
- [133] Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 2000;342:1645–50.
- [134] 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:e003342. doi:10.1136/bmjopen-2013-003342.
- [135] Cook DJ, Mulrow CD, Haynes RB. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Ann Intern Med* 1997;126:376–80. doi:10.7326/0003-4819-126-5-199703010-00006.
- [136] Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. *PLoS Med* 2012;9:e1001308. doi:10.1371/journal.pmed.1001308.
- [137] 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;32:4120–6. doi:10.1200/JCO.2014.56.7503.
- [138] Stephenson W. Correlating persons instead of tests. *J Pers* 1935;4:17–24. doi:10.1111/j.1467-6494.1935.tb02022.x.
- [139] Van Exel J, de Graaf G. Q methodology - A sneak preview -. *Soc Sci* 2005;2:1–30.

- [140] Brown SR. Q-Methodology and Quality Research. *Qual Health Res* 1996;561–7.
- [141] Dennis KE. Q methodology: relevance and application to nursing research. *Adv Nurs Sci* 1986;8:6–17.
- [142] Hackert C, Braehler G. FlashQ software [Computer software] 2007 [cited february 27th]; Available from : <http://www.hackert.biz/flashq/home/>.
- [143] Islamaj Dogan R, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. *Database (Oxford)* 2009;2009:bap018. doi:10.1093/database/bap018.
- [144] Fletcher RH, Black B. “Spin” in scientific writing: scientific mischief and legal jeopardy. *Med Law* 2007;26:511–25.
- [145] 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;32:4120–6. doi:10.1200/JCO.2014.56.7503.
- [146] Prasad V, Jorgenson J, Ioannidis JPA, Cifu A. Observational studies often make clinical practice recommendations: an empirical evaluation of authors’ attitudes. *J Clin Epidemiol* 2013;66:361–6.e4. doi:10.1016/j.jclinepi.2012.11.005.
- [147] Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009;302:2345–52. doi:10.1001/jama.2009.1757.
- [148] Ezzo J, Bausell B, Moerman DE, Berman B, Hadhazy V. Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *Int J Technol A H C* 2001-4-457-660000.pdf 2001;17:457–66.
- [149] Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006;333:597–600. doi:10.1136/bmj.333.7568.597.
- [150] Cook D, Guyatt G. Colloid use for fluid resuscitation: evidence and spin. *Ann Intern Med* 2001;135:205–8.
- [151] Weed DL. Meta-Analysis and Causal Inference: A Case Study of Benzene and Non-Hodgkin Lymphoma. *Ann Epidemiol* 2010;20:347–55. doi:10.1016/j.annepidem.2010.02.001.
- [152] Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol* 2000;29:387–90. doi:10.1093/ije/29.3.387.

- [153] Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857–65. doi:10.1016/j.jclinepi.2008.03.004.
- [154] Turner EH, Matthews AM, Linardatos E, Tell R a, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252–60. doi:10.1056/NEJMsa065779.
- [155] Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413–5. doi:10.1136/bmj.a117.
- [156] Langhorne P. Bias in meta-analysis detected by a simple, graphical test. Prospectively identified trials could be used for comparison with meta-analyses. *BMJ* 1998;316:471.
- [157] Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials* 2009;10:43. doi:10.1186/1745-6215-10-43.
- [158] Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525–37. doi:10.1002/sim.1185.
- [159] Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85. doi:10.1186/1745-6215-11-85.
- [160] Sterne JAC, Sutton AJ, Ioannidis JPA, 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.
- [161] Punjasawadwong Y, Phongchiewboon A, Bunchungmongkol N. Bispectral index for improving anaesthetic delivery and postoperative recovery. *Cochrane Database Syst Rev* 2014;6:CD003843. doi:10.1002/14651858.CD003843.pub3.
- [162] Wilkes MM, Navickis RJ. Patient Survival after Human Albumin Administration: A Meta-Analysis of Randomized, Controlled Trials. *Ann Intern Med* 2001;135:149–64.
- [163] Henrotin YE. Interpretation of metaanalyses: pitfalls should be more widely recognized. *J Rheumatol* 2012;39:1107–9. doi:10.3899/jrheum.111420.
- [164] Ioannidis JP a. Meta-research: The art of getting it wrong. *Res Synth Methods* 2010;1:169–84. doi:10.1002/jrsm.19.
- [165] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Med Decis Making* 2010;30:566–77. doi:10.1177/0272989X10375853.

- [166] Hróbjartsson A, Gøtzsche PC, Hrobjartsson A, Gotzsche PC. Powerful spin in the conclusion of Wampold et al.'s re-analysis of placebo versus no-treatment trials despite similar results as in original review. *J Clin Psychol* 2007;63:373–7. doi:10.1002/jclp.20357 [doi].
- [167] Sumner P, Vivian-Griffiths S, Boivin J, Williams a., Venetis C a., Davies a., et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *Bmj* 2014;349:g7015–g7015. doi:10.1136/bmj.g7015.
- [168] Chalmers I, Altman DG. Systematic Reviews. *BMJ Publ Gr* 1995;311:119.
- [169] Roehr B. “Spin” in media coverage of research can be traced to abstracts. *BMJ* 2012;345:e6106. doi:10.1136/bmj.e6106.
- [170] Levine D, Bankier AA, Kressel HY. Spin in radiology research: let the data speak for themselves. *Radiology* 2013;267:324–5. doi:10.1148/radiol.13130108.
- [171] Freshwater FM. Plastic surgery spin : Suppressing clarity. *Ournal Plast Reconstr Aesthetic Surg* 2015;68:292–4. doi:10.1016/j.bjps.2014.10.029.
- [172] Bohan Brown MM, Brown AW, Allison DB. Nutritional epidemiology in practice: learning from data or promulgating beliefs? *Am J Clin Nutr* 2013;97:5–6. doi:10.3945/ajcn.112.052472.
- [173] Selvaraj S, Borkar DS, Prasad V. Media coverage of medical journals: do the best articles make the news? *PLoS One* 2014;9:e85355. doi:10.1371/journal.pone.0085355.
- [174] McCartney M. Research press releases need better policing. *BMJ* 2014;348:g2868. doi:10.1136/bmj.g2868.
- [175] Wang MTM, Gamble G, Bolland MJ, Grey A. Press releases issued by supplements industry organisations and non-industry organisations in response to publication of clinical research findings: a case-control study. *PLoS One* 2014;9:e101533. doi:10.1371/journal.pone.0101533.
- [176] Van Atteveldt NM, van Aalderen-Smeets SI, Jacobi C, Ruigrok N. Media reporting of neuroscience depends on timing, topic and newspaper type. *PLoS One* 2014;9:e104780. doi:10.1371/journal.pone.0104780.
- [177] Robinson a, Coutinho a, Bryden a, McKee M. Analysis of health stories in daily newspapers in the UK. *Public Health* 2013;127:39–45. doi:10.1016/j.puhe.2012.10.001.
- [178] Gonon F, Konsman J-P, Cohen D, Boraud T. Why most biomedical findings echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. *PLoS One* 2012;7:e44275. doi:10.1371/journal.pone.0044275.
- [179] Weed DL. Preventing scientific misconduct. *Am J Public Health* 1998;88:125–9.

- [180] Hopewell S, Dutton S, Yu L-M, 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. doi:10.1136/bmj.c723.
- [181] Punjasawadwong Y, Phongchiewboon A, Bunchungmongkol N. Bispectral index for improving anaesthetic delivery and postoperative recovery (Review) Bispectral index for improving anaesthetic delivery and postoperative recovery. *Cochrane Database Syst Rev* 2014;6:10–2. doi:10.1002/14651858.CD003843.pub2.Copyright.
- [182] Santesso N, Rader T, Nilsen ES, Glenton C, Rosenbaum S, Ciapponi A, et al. A summary to communicate evidence from systematic reviews to the public improved understanding and accessibility of information: a randomized controlled trial. *J Clin Epidemiol* 2014. doi:10.1016/j.jclinepi.2014.04.009.
- [183] Bauchner H, Henry R, Golub RM. The restructuring of structured abstracts: adding a table in the results section. *JAMA* 2013;309:491–2. doi:10.1001/jama.2013.76.
- [184] Teno JM, Gozalo PL, Bynum JPW, Leland NE, Miller SC, Morden NE, et al. Change in end-of-life care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009. *JAMA* 2013;309:470–7. doi:10.1001/jama.2012.207624.

Annexes

Annexe 1 : Stratégies de recherche

To identify different types of spin we first searched Medline and the Cochrane methodology register.

In medline advanced searched with the following equation: ("Meta-Analysis as Topic"[Mesh]) AND (bias[MeSH Terms] OR bias[Title/Abstract] OR biased[TIAB] OR misconduct*[TIAB] OR mislead*[TIAB] OR misinterpretat*[TIAB] OR over-interpretat*[TIAB] OR under-interpretat*[TIAB] OR over-estimat*[TIAB] OR under-estimat*[TIAB] OR interpretat*[TIAB] OR influenc*[TIAB] OR distort*[TIAB] OR omission*[TIAB] OR omitted[TIAB] OR discord*[TIAB] OR altered[TIAB] OR altered[TIAB] OR altering[TIAB] OR error*[TIAB] OR erroneous*[TIAB] OR fail[TIAB] OR failure[TIAB] OR flaw*[TIAB] OR inappropriate[TIAB] OR irrelevant[TIAB] OR inadequate[TIAB] OR indue[TIAB] OR incomplete[TIAB] OR (low[TIAB] OR methodological[TIAB] OR "quality"[TIAB] OR inconsisten*[TIAB] OR discredit*[TIAB]) OR mistake*[TIAB]) that retrieved 3648 results on October 2011. From these results we filtered with the subset limit “core clinical journal” that retrieved 560 results.

In the Cochrane Methodology Register we did an advanced search with the key word: meta-analysis AND (misconduct OR biased OR mislead) in records title or abstract.

Annexe 2 : Mails d'invitation et copies d'écran du site de l'étude

Invitation e-mail

Dear colleague,

We are writing to ask for your help in making medical research more valuable to patients. Given your expertise in the field of clinical research, we would like to invite you to participate in an international academic study to investigate how physicians interpret abstracts.

The study has been designed so that it will require only a minimal amount of work on your part and yet provides a real opportunity to actively contribute to improve the communication of research results to inform clinical practice.

Your participation would involve in reading 1 abstract of a systematic review and answering 5 short questions about its findings.

As a thank for your participation in the study, you will be entered into a draw along with all other participants for a chance to win an **iPad mini**.

You can complete the survey by [clicking here](#).

Or by copying and pasting the following link into your web browser : "<http://www.epidemioteledieu.com/ILMARI2/php/survey.php?num=18&part=1018>"

With best wishes

Dr Amelie Yavchitz (INSERM, France)
Dr Sally Hopewell (University of Oxford, UK)
Pr Philippe Ravaud (Paris Descartes University, INSERM, France)
Pr Isabelle Boutron (Paris Descartes University, INSERM, France)

The University of Paris Descartes, Institutional Review Board (IORG0000725), has approved this study.
Your responses will be kept confidential.
If you prefer not to receive future reminders regarding this study, please contact Amelie Yavchitz at amelie.yavchitz@bjn.aphp.fr

Dear colleague,

A weeks ago we invited you to participate in a study which aims to evaluate the impact of abstract content on readers' interpretation.

We are writing to ask for your help in making medical research more valuable to patients. Given your expertise in the field of clinical research, we would like to invite you to participate in an international academic study to investigate how physicians interpret abstracts.

The study has been designed so that it will require only a minimal amount of work on your part and yet provides a real opportunity to actively contribute to improve the communication of research results to inform clinical practice.

Your participation would involve in reading 1 abstract of a systematic review and answering 5 short questions about its findings.

As a thank for your participation in the study, you will be entered into a draw along with all other participants for a chance to win an **iPad mini**.

You can complete the survey by [clicking here](#).

Or by copying and pasting the following link into your web browser : "<http://www.epidemiology-hoteldieu.com/ILMARI2/php/survey.php?num=3845&part=4845>"

With best wishes

Dr Amelie Yavchitz (INSERM, France)
Dr Sally Hopewell (University of Oxford, UK)
Pr Philippe Ravaud (Paris Descartes University, INSERM, France)
Pr Isabelle Boutron (Paris Descartes University, INSERM, France)

The University of Paris Descartes, Institutional Review Board (IORG0000725), has approved this study.
Your responses will be kept confidential.
If you prefer not to receive future reminders regarding this study, please contact Amelie Yavchitz at amelie.yavchitz@bjn.aphp.fr

Survey

Based on the information reported in the abstract, please answer the following questions about the intervention. To avoid any biased interpretation, the description and name of the intervention have been masked.

Randomized abstract

1) How confident are you in the results of this study?

0 1 2 3 4 5 6 7 8 9 10

« Not confident at all»
confident»

« very

2) How confident are you in the validity of the conclusions of this study?

0 1 2 3 4 5 6 7 8 9 10

« Not confident at all »

« very confident»

3) How confident are you that the intervention “A” could be of benefit to patients?

0 1 2 3 4 5 6 7 8 9 10

« Not confident at all»

« very confident»

4) How confident are you that the results of this study could influence your clinical practice?

0 1 2 3 4 5 6 7 8 9 10

« Not confident at all »

« very confident»

5) Do you think that this systematic review was conducted rigorously?

0 1 2 3 4 5 6 7 8 9 10

« not rigorously at all»

« yes, very rigorously»

Page 2

Other data collected

Demographic data

Your age :

Sex : Female Male

Expertise :

Qualification : MD, MD/PhD, Other

How many years have you been a clinician ? <5, 5-10, 10-15, >15

Where are you currently located : France, UK, Other European country, USA, Canada, South america, Asia, Oceania.

Are you working in a ? Primary care center, Secondary care center, Tertiary care center.

How many randomized trial have you been involved in (as principal or co-investigator) ? 1-3, 4-6, 7-10, >10

How many systematic review have you been involved in (as first or co-authors) ? 0, 1-3, 4-6, 7-10, >10

In the last month, approximately how many abstracts published in biomedical journals have you read ? 0, 1-5, 5-10, 10-15, 15-20, >20

In the last year, approximately how many articles have you peer reviewed for a biomedical journal ? 0, 1-5, 5-10, 10-15, 15-20, >20

In the last year, approximately how many grant proposals have you peer reviewed ? 0, 1-5, 5-10, 10-15, 15-20, >20

Did you received some training in Clinical Epidemiology ? Yes No

Did you received some training related to the methods of systematic review ? Yes No

Page 3

Thank you for your participation in our survey

Annexe 3 : Mail d'invitation au Q-Sort

Dear “first name” “Family name”,

We are contacting you knowing your expertise in systematic reviews and meta-analysis to invite you to participate in a study on “spin” in abstract.

We are conducting this study in collaboration with the other co-convenors of the Bias Methods Group (Doug Altman, Asbjørn Hrobjartsson, David Moher) and Toby Lasserson from the Cochrane Editorial Unit.

Spin is defined as a specific way of reporting (deliberate or not) to convince the reader that the beneficial effect of the experimental treatment is greater than that shown by the results.

Based on a systematic review, we identified 21 different types of spin that could occur in abstracts of systematic reviews. Our objective is to classify the different types of spin according to their severity (i.e., likelihood of distorting readers' interpretation).

For this purpose, we are using the Q-sort methodology. In a Q-sort, we ask you to sort the 21 different types of spin. For this purpose, we ask you to 1) place the spin types in order of severity according to your opinion and then 2) sort them relative to each other in terms of severity on a score sheet representing a quasi-normal distribution ranging from -3 (less severe) to +3 (more severe). This is achieved in 4 steps through user-friendly online software. The task should take about 15 minutes of your time. Your answers will be treated confidentially.

Unfortunately the Q-sorting software does not allow you to save and finish later, so the sorting has to be done in one sitting. Furthermore, the online Q-Sort runs with Flash Player and please use the most recent version for your browser.

To complete the survey, please click here, or copy and paste the following link into your browser: http://www.qsort.fr/SPIN_Cochrane/Flashq/

If you have any questions or need technical help, please email me at amelie.yavchitz@cochrane.fr

We very much appreciate your participation in this project and we look forward to your answer.

Yours faithfully,

Annexe 4: Répartition des réponses pour chaque item

