

BMJ Open How do authors of systematic reviews deal with research malpractice and misconduct in original studies? A cross-sectional analysis of systematic reviews and survey of their authors

Nadia Elia,^{1,2} Erik von Elm,³ Alexandra Chatagner,⁴ Daniel M Pöpping,⁵ Martin R Tramèr^{1,6}

To cite: Elia N, von Elm E, Chatagner A, *et al*. How do authors of systematic reviews deal with research malpractice and misconduct in original studies? A cross-sectional analysis of systematic reviews and survey of their authors. *BMJ Open* 2016;**6**:e010442. doi:10.1136/bmjopen-2015-010442

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-010442>).

Received 4 November 2015
Revised 12 January 2016
Accepted 10 February 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Nadia Elia;
nadia.elia@hcuge.ch

ABSTRACT

Objectives: To study whether systematic reviewers apply procedures to counter-balance some common forms of research malpractice such as not publishing completed research, duplicate publications, or selective reporting of outcomes, and to see whether they identify and report misconduct.

Design: Cross-sectional analysis of systematic reviews and survey of their authors.

Participants: 118 systematic reviews published in four journals (*Ann Int Med*, *BMJ*, *JAMA*, *Lancet*), and the *Cochrane Library*, in 2013.

Main outcomes and measures: Number (%) of reviews that applied procedures to reduce the impact of: (1) publication bias (through searching of unpublished trials), (2) selective outcome reporting (by contacting the authors of the original studies), (3) duplicate publications, (4) sponsors' and (5) authors' conflicts of interest, on the conclusions of the review, and (6) looked for ethical approval of the studies. Number (%) of reviewers who suspected misconduct are reported. The procedures applied were compared across journals.

Results: 80 (68%) reviewers confirmed their data. 59 (50%) reviews applied three or more procedures; 11 (9%) applied none. Unpublished trials were searched in 79 (66%) reviews. Authors of original studies were contacted in 73 (62%). Duplicate publications were searched in 81 (69%). 27 reviews (23%) reported sponsors of the included studies; 6 (5%) analysed their impact on the conclusions of the review. Five reviews (4%) looked at conflicts of interest of study authors; none of them analysed their impact. Three reviews (2.5%) looked at ethical approval of the studies. Seven reviews (6%) suspected misconduct; only 2 (2%) reported it explicitly. Procedures applied differed across the journals.

Conclusions: Only half of the systematic reviews applied three or more of the six procedures examined. Sponsors, conflicts of interest of authors and ethical approval remain overlooked. Research misconduct is sometimes identified, but rarely reported. Guidance on when, and how, to report suspected misconduct is needed.

Strengths and limitations of this study

- This study combines quantitative and qualitative methods to investigate how systematic reviewers deal with research malpractice and misconduct.
- It proposes clear and reproducible samples of systematic reviews from four major medical journals and the *Cochrane Library*.
- The extracted data were confirmed by 70% of the authors of the systematic reviews analysed.
- The systematic reviewers were not asked to confirm all their data, but only the information considered ambiguous.
- There is currently no common definition of 'research misconduct'. This may have led to an underestimation of its real prevalence.

INTRODUCTION

Rationale

Research misconduct can have devastating consequences for public health¹ and patient care.^{2 3} While a common definition of research misconduct is still lacking, there is an urgent need to come up with strategies to prevent it.^{4 5} Fifteen years ago, Smith⁶ proposed 'a preliminary taxonomy of research misconduct' describing 15 practices ranging from 'minor' to 'major' misconduct. Some of these practices, however, are very common and may not be regarded by all as 'misconduct'. Therefore, this research defines 'malpractice' as relatively common and minor misconduct, while the term 'misconduct' is used for data fabrication, falsification, plagiarism or any other intentional malpractices.

It has been shown that some of the 15 malpractices described by Smith, threaten the conclusions of systematic reviews. Examples of such malpractices include: avoiding the publication of a completed research,^{7 8}

duplicate publications,⁹ selectively reporting on outcomes or adverse effects,¹⁰ and presenting biased results that are in favour of the sponsors¹¹ or the authors¹² interests. The impact of other malpractices, such as gift or ghost authorship, is less clear.

Rigorous systematic review methodology includes specific procedures that can counter-balance some of the research malpractice. Unpublished studies may, for example, be identified through exhaustive literature searches,¹³ and statistical tests or graphical displays such as funnel plots can quantify the risk of publication bias.¹⁴ Unreported outcomes may be unearthed by contacting the authors of original articles, and multiple publications based on the same cohort of patients can be identified and excluded from analyses.¹⁵ Authors of systematic reviews (further called: *systematic reviewers* or *reviewers*) can also use sensitivity analyses to quantify the impact sponsors' and authors' personal interests have on the conclusions of a review. Finally, it has been suggested that, as part of the process of systematic reviewing, ethical approval of included studies or trials should be checked on in order to identify unethical research.^{16 17} Systematic reviewers could hence act as whistle-blowers when reporting any suspected misconduct.¹⁸

Objectives

The aim of this study is to examine whether systematic reviewers apply the aforementioned procedures, and whether they uncover and report on cases of misconduct.

The study first examines whether reviewers searched for unpublished studies or tested for publication bias, contacted authors to unearth unreported outcomes, searched for duplicate publications, analysed the impact of sponsors or possible conflicts of interest of study authors, checked on ethical approval of the studies and reported on misconduct. The secondary objective was to examine whether four major journals and the Cochrane Library reported consistently on the issue.

METHODS

The reporting of this cross-sectional study follows the STROBE recommendation.¹⁹ The protocol is available from the authors.

Study design

We conducted a cross-sectional analysis of systematic reviews published in 2013 in four general medical journals (*Annals of Internal Medicine* (Ann Int Med), *The BMJ* (BMJ), *JAMA* and *The Lancet* (Lancet)). A random sample of new reviews was drawn from the *Cochrane Database of Systematic Reviews* (*Cochrane Library*) in 2013, as Cochrane reviews are considered the gold standard in terms of systematic reviewing.

Setting and selection of systematic reviews

Systematic reviews were identified through a PubMed search in August 2014, using the syntax 'systematic review [Title] AND journal title [Journal], limit 01.01.2013 to 31.12.2013'. A computer-generated random sequence was used to select 25 reviews published in 2013 in the *Cochrane Library* (<http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/2013-table-of-contents.html>).

Reviews were selected by one of the five authors (NE) on the basis of the review titles and abstracts. This was checked by another author (AC). To be eligible, reviews had to describe a literature search strategy and include at least one trial or study. Narrative reviews or meta-analyses without an exhaustive literature search were not considered.

Variables

From each systematic review, we extracted the following information: the first author's name and country of affiliation; the number of co-authors; the name of the journal; the title of the review; the number of databases searched; the number of studies and study designs included; the language limitations applied; whether or not a protocol was registered and freely accessible; and, finally, sources of funding and possible conflicts of interest of the reviewers.

Furthermore, we examined whether each of the selected reviews applied the following six procedures, they: (1) searched for unpublished trials; (2) contacted authors to identify unreported outcomes; (3) searched for duplicate publications (defined as a redundant republication of an already published study, with or without a cross-reference to the original article); (4) analysed the impact of the sponsors of the original studies on the conclusions of the review; (5) analysed the impact of possible conflicts of interest of the authors on the conclusions of the review; and (6) extracted information on ethical approval of included studies. We used the following rating system: 0=procedure not applied, 1=partially applied, 2=fully applied (table 1).

Finally, we collected information on whether the systematic reviewers suspected, and explicitly reported on, any misconduct in the included articles.

Bias

Data from the reviews were extracted by one author (NE), and copied into a specifically designed spreadsheet. Two of the co-authors checked the data (AC and DMP). We contacted all the corresponding authors of the reviews and asked them to confirm our interpretation of their methods of review. This included their method regarding the search for unpublished trials, their contacts with the authors, their search for duplicate publications and their identification of misconduct. When there was discrepancy between our interpretation and the reviewers' answers, we used the latter. This was

Table 1 Rating of the six procedures examined

	Score		
	0	1	2
Search of unpublished trials and/or test for publication bias	Unpublished trials not searched, publication bias not tested	Unpublished trials searched OR publication bias tested	Unpublished trials searched AND publication bias tested or statistically corrected
Contact with study authors	Study authors not contacted	Study authors contacted for methodology or unspecified reasons	Study authors contacted to unearth unreported endpoints
Duplicate publications	Duplicate publications not searched or not mentioned	Duplicate publications searched	Duplicate publications referenced in the published report*
Sponsors of the studies	Information on study sponsors not reported	Information on study sponsors reported	Impact of study sponsor(s) on the conclusions of the review analysed
Study authors' conflicts of interest	Study authors' conflicts of interest not reported	Study authors' conflicts of interest reported	Impact of study authors' conflicts of interest on the conclusions of the review analysed
Ethical approval	Ethical approval of included studies not reported	Ethical approval of included studies reported	Lack of ethical approval of included studies reported and referenced†

Percentages may not add-up to 100% because of rounding errors.

*Also includes reports that explicitly mention that no duplicates were identified.

†Also includes reports that explicitly mention that none of the included studies lacked ethical approval.

done by email, and a reminder was sent to those who had not replied within 2 weeks.

Sample size

The capacity of systematic reviewers to identify misconduct is unknown. Our hypothesis was that 5% of systematic reviewers would identify misconduct. Therefore, we needed a minimum of 110 systematic reviews to allow us to detect a prevalence of 5%, if it existed, with a margin of error of 4% assuming an α -error of 0.05.

Statistical methods

Descriptive results are reported as numbers (proportions) and median (IQR) as required. To check whether systematic reviews were different from one journal to the other, we performed all descriptive analyses separately according to title of the journal. χ^2 or Kruskal–Wallis tests were applied to test the null hypothesis of homogeneous distribution of characteristics and outcomes. We compared reviews from reviewers who answered our inquiry with reviews from those who did not, and across journals. Since Cochrane reviews were expected to be different from those published in the journals, we performed separate analyses with and without Cochrane reviews. We did not expect missing data. Statistical significance was defined as an α -error of 0.05 or less in two-sided tests. Analyses were performed using STATA V.13.

RESULTS

Selection of reviews

We identified 136 references; 18 were excluded for different reasons, leaving us with 118 systematic reviews (*Ann Int Med* 39^{A1-39}; *BMJ* 38^{B1-38}; *JAMA* 12^{J1-12}; *Lancet*

10^{L1-10}; *Cochrane Library* 19^{C1-19}) (figure 1, online supplementary appendix table 1A).

Characteristics of the reviews

The characteristics of the reviews are described in table 2, online supplementary appendix tables 1A and 2A. Approximately 75% of the first authors were affiliated to an English-speaking institution. The protocols of all the Cochrane reviews were registered and available. However, protocols were available for only 17 reviews from the journals.

Sources of funding were declared in 110 reviews. Among these 110, 24 declared that they had no funding at all. All the reviews declared presence or absence of conflicts of interest of the reviewers.

The median number of databases searched was four. Additional references were searched in systematic reviews published previously and through contacting experts and/or authors of the original studies. Forty-two (36%) reviews only considered the English literature, and 6 (5%) reviews searched in Medline only. Four (3%) reviews searched for English articles in Medline only.^{A(35,39)J(5,9)} The median number of articles included per review was 28. Half of the systematic reviews included a mix of various study designs, while 39% included only RCTs (table 2).

Outcome data

Contact with reviewers

Out of the 118 reviews, we were able to contact 111 corresponding authors. No valid email address was available for seven. Eighty reviewers (72%) responded to our inquiries.

Among the 80 reviewers who responded, 8 (10%) provided information that changed our data extraction

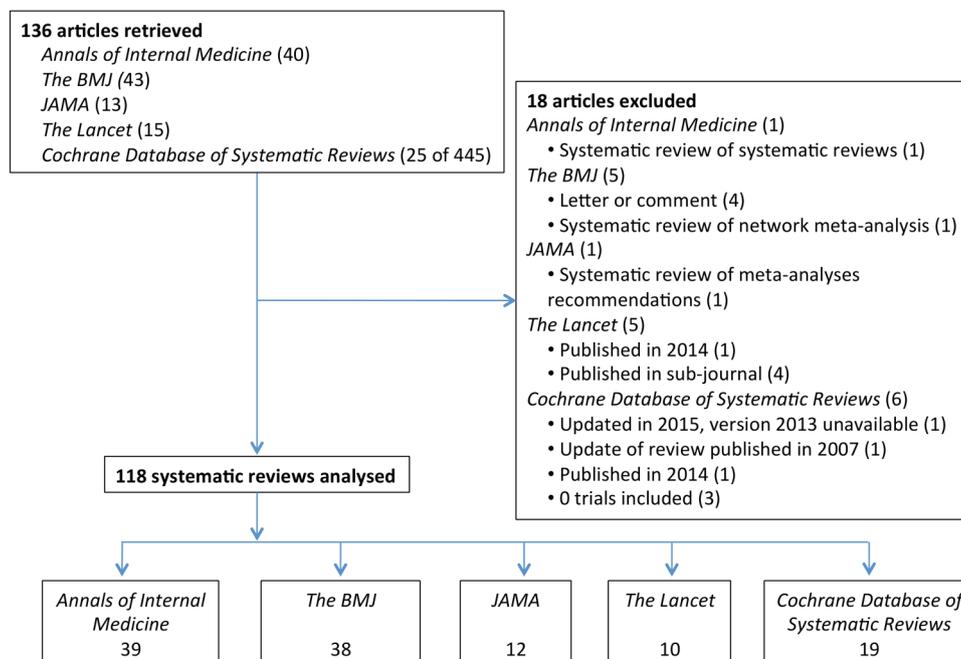


Figure 1 Flowchart of retrieved and analysed systemic reviews.

regarding the endpoint ‘search for unpublished trials or test for publication bias’. One reviewer declared that, contrary to our assumption, unpublished trials had not been searched in their review,^{B28} and seven claimed that unpublished trials had been searched although this was not reported in the published reviews.^{A(21,25,30,31),B(2,17,27)}

Eleven reviewers (14%) provided information that changed our data extraction regarding the endpoint ‘contact with authors of original studies’. One declared that authors had not been contacted,^{B35} and 10 claimed that authors had been contacted although this was not reported in the published review.^{A(12,25,31,34),B(20,25,34),J(9,12),L1}

Twenty-six reviewers (32%) provided information that changed our data extraction regarding the endpoint ‘duplicate publication’. Terms used included ‘duplicate’, ‘companion article’, ‘multiple publications’, ‘articles with overlapping datasets’, or ‘trials with identical patient population’. Three reviewers declared that, contrary to our assumption, they had not identified duplicate publications,^{A(25,26),J3} and 23 claimed having searched for duplicates although this was not reported in the published review.^{A(16,17,21,24,27,30,31,33),B(9,10,14,20,25,27),C(6,10,11,14,16,18),J6,L(6,9)}

Five reviewers (6%) told us about suspected cases of misconduct that were not reported in the published review.

Characteristics of the reviews did not differ depending on whether or not we were able to contact their authors (table 2).

Main results

The median number of procedures applied in each review was 2.5 (IQR, 1–3). Eleven reviews (9%) applied no procedures at all, while no review applied all six procedures.

Search of unpublished trials and test for publication bias

Fifty-six reviewers (47%) either searched for unpublished trials or applied a statistical test to identify publication bias. Twenty-three reviewers (19%) did both. Unpublished studies were sought for in trial registries (eg, ClinicalTrials.gov or FDA database), or by contacting experts and manufacturers. The number of unpublished studies included in these reviews was inconsistently reported. Contacting the reviewers did not help us clarify this issue. Seven reviews (6%) only discussed the risk of publication bias, and 32 reviews (27%) did not mention it at all (table 3).

Contact with authors to unearth unreported outcomes

Seventy-three reviewers (62%) had contacted the authors of the original studies. Fifty-eight reviewers (49%) had searched for unreported results from the original articles. The reviews rarely reported on the number of authors contacted and the response rate. We were not able to clarify this issue in our email exchange with the reviewers (table 3).

Duplicate publications

Duplicate publications were sought for in 81 reviews (69%). Twenty-two reviewers confirmed that duplicates were not sought for, and 15 did not answer our enquiry. The number of duplicates identified was rarely mentioned. We failed to clarify this issue in our exchange with the reviewers. Ten reviews (8.5%) published the reference of at least one identified duplicate.^{A(1,10,18,38),C(1,4,6,12),J(2,9)}

Sponsors

Twenty-seven reviews (23%) reported on the sources of funding for the studies. Six reviewers (5%) analysed the

Table 2 Characteristics of the systematic reviews analysed

	ALL		Answer		No answer		p Value
	N	Per cent	N	Per cent	N	Per cent	
Number of systematic reviews	118	100	80	68	38	32	
Number of co-authors							0.240
Median (IQR)	6 (5–9)		6 (4–7.5)		6 (5–9)		
Country of affiliation of 1st author							0.969
USA	42	36	27	34	15	39	
UK	23	19	16	20	7	18	
Canada	21	18	14	18	7	18	
Australia	4	3	3	4	1	3	
Others	28	24	20	25	8	21	
Protocol registration/accessibility							0.513
No mention of protocol	82	69	53	66	29	76	
Protocol exists/not freely accessible	3	3	2	3	1	3	
Protocol exists/freely accessible	33	28	25	31	8	21	
Source of funding of systematic review							0.420
Not reported	8	7	7	9	1	3	
Reported that no funding was received	24	20	15	19	9	24	
Reported that funding was received with details	86	73	58	73	28	74	
Conflicts of interests of systematic reviewers							0.130
Not reported	0	0	0	0	0	0	
Reported that there were none	42	36	33	41	9	24	
Reported and detailed in the published report	61	52	39	49	22	58	
Reported elsewhere (eg, online)	15	13	8	10	7	18	
Number of databases searched							0.637
Median (IQR)	4 (3–7)		4 (3–7)		4 (3–6)		
Language limitations applied							0.569
None	56	47	35	44	21	55	
English only	42	36	30	38	12	32	
Limitations to other languages	7	6	6	8	1	3	
Not specified	13	11	9	11	4	11	
Number of studies included							0.804
Median (IQR)	28 (12–57)		28 (12–59)		34 (12–55)		
Study designs examined							0.312
RCT only	46	39	32	40	14	37	
Cohort prospective only	5	4	4	5	1	3	
Diagnostic studies only	5	4	5	6	0	0	
Various designs	62	53	39	49	23	61	

Answer: reviews in which extracted data were confirmed by reviewers. No answer: reviews in which extracted data were not confirmed by reviewers. p Value testing the null hypothesis of equal distribution between the reviews for which the author responded to our inquiries and those who did not. Statistical tests: χ^2 test or Kruskal-Wallis equality of population rank tests, as appropriate. RCT, randomised controlled trials.

impact of sponsors on the results of the review.^{B(4,15,26,32)J12,L8} One reviewer claimed that sponsor bias was unlikely,^{L8} while three were unable to identify any sponsor bias.^{B(4,15)J12} Finally, two reviews identified sponsor bias (see online supplementary appendix table 3A).^{B26,32}

Conflicts of interest of authors

Five reviewers reported on conflicts of interest of the authors of the studies.^{A(8,14),B20,C(8,14)} None of them used this information to perform subgroup analyses. One review mentioned conflicts of interest as a possible explanation for their (biased) findings.^{A8} In three reviews, the affiliations of the authors were summarised and potential conflicts of interest clearly identified.^{B20,C}

^(8,14) Finally, the appendix table of one of the reviews showed that one study might have suffered from a 'significant conflict of interest'.^{A4}

Ethical approval of the studies

Three reviews looked at whether or not ethical approval had been sought for (see online supplementary appendix table 3A).^{B(27,37),C10} Two reviews explicitly reported that all included studies had received ethical approval.^{B(27–37)} The third review reported extensively on which studies had or had not provided any information on ethical approval or patient consent.^{C10}

Outcomes did not differ according to whether or not we were able to contact the reviewers (table 3).

Table 3 Application of the procedures to counter-balance some common research malpractices

	ALL		Answer		No answer		p Value
	N	Per cent	N	Per cent	N	Per cent	
Number of systematic reviews	118	100	80	68	38	32	
Search of unpublished trials and/or test for publication bias							0.914
Publication bias discussed only or not mentioned	39	33	26	33	13	34	
Unpublished trials searched OR publication bias tested	56	47	39	49	17	45	
Unpublished trials searched AND publication bias tested	23	19	15	19	8	21	
Contact with authors of the studies							0.427
Study authors not contacted	45	38	28	35	17	45	
Study authors contacted for method or unspecified reason	15	13	12	15	3	8	
Study authors contacted for unreported outcomes	58	49	40	50	18	47	
Duplicate publications							0.057
Not searched or not mentioned	37	31	21	26	16	42	
Searched and found, not referenced OR no mention of results	71	60	54	68	17	45	
Searched, found and referenced	10	8	5	6	5	13	
Sponsors of the studies							0.809
Not mentioned	91	77	63	79	28	74	
Information extracted	21	18	13	16	8	21	
Information extracted and subgroup analyses performed	6	5	4	5	2	5	
Conflicts of interests of study authors							0.703
Not mentioned	113	96	77	96	36	95	
Information extracted	5	4	3	4	2	5	
Information extracted and subgroup analyses performed	0	0	0	0	0	0	
Ethical approval of included studies							0.481
Not mentioned	115	97	77	96	38	100	
Information extracted	3	3	3	4	0	0	
Information extracted and subgroup analyses performed	0	0	0	0	0	0	
Number of procedures applied							0.403
None	11	9	6	8	5	13	
1 or 2 procedures	48	41	34	43	14	37	
3 or 4 procedures	56	47	38	48	18	47	
5 procedures	3	3	2	3	1	3	
Median (IQR)	2.5 (1–3)		2.5 (2–3)		2.5 (1–3)		
Explicit mention of misconduct by reviewers							0.296
No, or not mentioned	111	94	74	93	37	97	
Yes	7	6	6	8	1	3	

Answer: reviews in which extracted data were confirmed by reviewers. No answer: reviews in which extracted data were not confirmed by reviewers. p Value testing the null hypothesis of equal distribution between the reviews for which the authors responded to our inquiry and those who did not (χ^2 test). Percentages may not add-up to 100% because of rounding errors.

Suspicion of misconduct

Two reviewers suspected research misconduct in the articles included in their review and reported it accordingly.^{B31,J12} Contacting the other reviewers allowed us to uncover five additional cases of possible misconduct. Four agreed to be cited here;^{A26,B33,C16,L1} while one preferred to remain anonymous.

Data falsification was suspected in three reviews.^{A26,C16,J12} One review, looking at the association of hydroxyethyl starch administration with mortality or acute kidney injury of critically ill patients,^{J12} included seven articles co-authored by Joachim Boldt. However, a survey performed in 2010, and focusing on Boldt's research published between 1999 and 2010, had led to the retraction of 80 of his articles due to data fabrication and lack of ethical approval.²⁰ The seven articles co-authored by Boldt were kept in the review as they had been published *before* 1999. Nonetheless, the reviewers performed sensitivity analyses excluding these seven articles, and

showed a significant increase in the risk of mortality and acute kidney injury with hydroxyethyl starch solutions that was not apparent in Boldt's articles.

The second review examined different techniques of sperm selection for assisted reproduction. The reviewers suspected data manipulation in one study since its authors reported non-significant differences between the number of oocytes retrieved and embryos transferred while the p-value, when recalculated by the reviewers, was statistically significant.^{C16}

The third review (on management strategies for asymptomatic carotid stenosis) reported that misconduct had been suspected based on the 'differences in data between the published SAPPGIRE trial and the re-analysed data posted on the FDA website'.^{A26} Although this information was not provided in the published review, it was available in the full report online (<http://www.ahrq.gov/research/findings/ta/carotidstenosis/carotidstenosis.pdf>).

Intentional selective reporting of outcomes was suspected in two reviews.^{B31} In one review that examined early interventions to prevent psychosis, the reviewers identified, discussed and referenced three articles that did not report on all the outcomes.^{B31} In the second review, the corresponding reviewer (who preferred to remain anonymous) revealed that he 'knew of two situations in which authors knew what the results showed if they used standard categories (of outcome) but did not publish them because there was no relationship, or not the one they had hoped to find.'

Plagiarism was identified in one review examining the epidemiology of Alzheimer's disease and other forms of dementia in China.^{L1} According to the reviewers, they had identified a 'copy-paste-like duplicate publication of the same paper published two or more times, with the same results and sometimes even different authors'. This information was not reported in the published review. The corresponding reviewer explained that they 'had a brief discussion...about what to do about those findings and whether to mention them in the paper...We did not think that we should be distracted from our main goal, so we felt that it was better to leave it to qualified bodies and specialised committees on research malpractice to address this problem separately.'

Finally, one reviewer told us that "there were some 'suspected misconduct' in original studies on which we performed sensitivity analysis (best case—worst case)". The review mentioned that some studies were of poor quality, but it did not specifically mention suspicion of misconduct.^{B33}

The median number of studies included in the reviews that detected misconduct was 56 (IQR, 11–97), and was 28 (IQR, 12–57) in the reviews that did not detect misconduct. The difference did not reach statistical significance.

Secondary endpoints

The reviews published in the four medical journals differed in most characteristics examined in this paper. The reviews published in the *Cochrane Library* differed from all the other reviews (see online supplementary appendix table 2A). There were also differences in procedures applied across the journals (see online supplementary appendix table 4A). The only three reviews that had extracted data on ethical approval for the studies were published in the *BMJ* (2) and in the *Cochrane Library* (1). Finally, one review from the *BMJ* and one review from *JAMA* explicitly mentioned potential misconduct.

DISCUSSION

Statement of principal findings

This analysis confirms some issues and highlights new ones. The risk related to double counting of participants due to duplicate publications and the risk of selective reporting of outcomes are reasonably well recognised.

More than half of the reviews applied procedures to reduce the impact of these malpractices. The problem of conflicts of interest remains underestimated, and ethical approval of the original studies is overlooked. Although systematic reviewers are in a privileged position to unearth misconduct such as copy-paste-like plagiarism, intentional selective data reporting and data fabrication or falsification, they do not systematically report them. Finally, editors have a role to play in improving and implementing rigorous procedures for the reporting of systematic reviews to counter-balance the impact of research malpractice.

Comparison with other similar analyses

Our study confirms that systematic reviewers are able to identify publications dealing with the same cohort of patients.¹⁵ However, 20% of reviews under consideration failed to report having searched for duplicates. Only ten of them provided the references to some of the identified duplicates. It remains unclear whether reviewers do not consider duplicate publication worth disclosing or whether they are unsure on how to address the issue. Finally, there is no widely accepted definition of the term 'duplicate', which, in turn, adds to the confusion. For example, a number of reviewers used the term 'duplicate' to describe identical references identified more than once through the search process.

Selective publication of studies, and selective reporting of outcomes, have been examined previously.^{21–26} This led the *BMJ* to call for « *publishing yet unpublished completed trials and/or correcting or republishing misreported trials* ». ²⁷ Other ways to address this issue include registration of the study protocols,²⁸ searching for unpublished trials^{8 29} and contacting authors to retrieve unreported outcomes.^{30 31} Our analyses show that 70% of systematic reviewers are aware of these malpractices, although 10% failed to report them explicitly. As described before,³² most reviewers did not report on the number of unpublished articles included in their analyses, the number of authors contacted and the response rate.

Despite the obvious risk of research conclusions favouring a sponsor,³³ subgroup analyses on funding were rarely performed. Sponsor bias may overlap with other malpractices such as selective reporting,^{34 35} redundant publication⁹ or failure to publish completed research.³⁵ It may also overlap with conflicts of interest of the authors of the original studies, an issue that remains largely overlooked in these systematic reviews. There is a general understanding that authors with conflicts of interest are likely to present conclusions in favour of their own interests.^{12 36} Although most journals now ask for a complete and overt declaration of conflicts of interest from all authors, this crucial information remains unclearly reported. This may explain why we found no reviews that performed subgroup analyses on this issue.

Ten years ago, Weingarten proposed that ethical approval of studies should be checked during the

process of systematic reviewing.¹⁶ However, our study shows that only three reviews reported having done so. The need to report ethical approval in original studies has only recently been highlighted. A case of massive fraud in anaesthesiology,²⁰ in which informed consent from patients and formal approval by ethics committee were fabricated, only shows how difficult a task it will be.

The most striking finding was that although seven systematic reviews suspected misconduct in original studies, five of them did not report it, one reported it without further comment and only one reported overtly on the suspicion. This illustrates that reviewers do not consider themselves entitled to make an allegation, although they are in a privileged position to identify misconduct. The fact that one reviewer preferred to remain anonymous further illustrates the reluctance to openly report on misconduct.

Strengths and weaknesses

We used a clear and reproducible sampling method that was not limited to any medical specialty. The data analysed had been confirmed by the reviewers. This allowed us to quantify the proportion of procedures that were implemented but not reported by the reviewers. Finally, to our knowledge, this is the first analysis of the procedures used by systematic reviewers to deal with research malpractice and misconduct. Qualitative answers of the reviewers were very informative.

We selected systematic reviews from four major medical journals and the *Cochrane Library*, which is considered the gold standard in terms of systematic reviewing. We can reasonably assume that the problems identified are at least as serious in other medical journals. Systematic reviews that were not identified as such in their titles were not included. However, including these reviews would not have changed our findings. Only one reminder was sent to the reviewers, since the response rate was reasonably high. Furthermore, the characteristics of the systematic reviews did not differ between reviews for which authors responded or not. We did not ask the reviewers to confirm all their data but focused on the information that we considered unclear. It is possible that some of the reviewers had indeed extracted information on sponsors' and authors' conflicts of interest, as well as ethical approval of the studies, but failed to report them. A number of procedures applied and instances of misconduct came to light through our personal contacts with the reviewers. Our results might therefore be underestimated. On the other hand, it is possible that some reviewers *pretended* having applied some procedures although they had not. This would have led to an overestimation of the number of procedures applied. The major weakness of this study lies in the lack of an accepted definition of 'research misconduct'. It is possible that some reviewers might have hesitated to disclose suspected misconduct, leading to an underestimation of the prevalence of misconduct identified. Finally, we have used the 'preliminary

taxonomy of research misconduct' proposed by Smith in 2000,⁶ and categorised all common minor misconduct as 'malpractices'. Some may disagree with our classification.

Conclusions and research agenda

The PRISMA guideline has improved the reporting of systematic reviews.³⁷ PRISMA-P aims to improve the robustness of the protocols of systematic reviews.³⁸ The 17-items list mentions the assessment of meta-bias(es), such as publication bias across studies, and selective outcome reporting within studies. However, the list is not concerned with authors' conflicts of interest, sponsors, ethical approval of original studies, duplicate publications or the reporting of suspected misconduct. The MECIR project defines 118 criteria classified as 'mandatory' or 'highly desirable', to ensure transparent reporting in a Cochrane review.³⁹ Among these criteria, publication and outcome reporting bias, funding sources as well as conflicts of interest are highlighted as 'mandatory' to report on. However, ethical approval for studies is not. Most importantly, neither of the two recommendations explicitly describes what should be considered misconduct, what kind of misconduct must be reported, to whom, and how.

We have previously shown how systematic reviewers can test the impact of fraudulent data on systematic reviews,⁴⁰ identify redundant research⁴¹ and identify references that should have been retracted.⁴² This paper suggests that systematic reviewers may have additional roles to play. They may want to apply specific procedures to protect their analyses from common malpractices in the original research, and they may want to identify and report on suspected misconduct. However, they do not seem to be ready to act as whistle-blowers. The need for explicit guidelines on what reviewers should do once misconduct has been suspected or identified has already been highlighted.¹⁸ These guidelines remain to be defined and implemented. The proper procedure would require the reviewer to request the institution where the research was conducted to investigate on the suspected misconduct, as the institution holds the legal legitimacy. Whether alternative procedures could be applied should be discussed. For example, they may include contacting the editor-in-chief of the journal where the suspected paper was originally published, or the editor-in-chief where the systematic review will eventually be published. Future research should explore the application of additional protective procedures such as checking for the adherence of each study to its protocol, or the handling of outlier results, and quantify the impact of these measures on the conclusions of the reviews. Finally, potential risks of false reporting of misconduct need to be studied.

Author affiliations

¹Division of Anaesthesiology, Geneva University Hospitals, Geneva, Switzerland

²Institute of Global Health, University of Geneva, Geneva, Switzerland

³Cochrane Switzerland, Institute of Social and Preventive Medicine, Lausanne, Switzerland

⁴Institute Guillaume Belluard, Cran Gevrier, France

⁵Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany

⁶Faculty of Medicine, University of Geneva, Geneva, Switzerland

Acknowledgements The authors thank all the authors of the systematic reviews who kindly answered our inquiries, and Liz Wager for her thoughtful advice on the first draft of our manuscript.

Contributors NE, MRT and EvE designed and coordinated the investigations. NE wrote the first draft of the manuscript. AC and DMP were responsible for checking the inclusion and extraction of the data. NE performed the statistical analyses and contacted the reviewers. All the authors contributed to the revision of the manuscript and the intellectual development of the paper. NE is the study guarantor.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests All the authors have completed the Unified Competing Interests form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that they have received no support from any organisation for the submitted work, that they have no financial or non-financial interests that may be relevant to the submitted work, nor do their spouses, partners, or children. EvE is the co-director of Cochrane Switzerland—a Cochrane entity involved in producing, disseminating and promoting Cochrane reviews.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

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

REFERENCES

- Deer B. How the case against the MMR vaccine was fixed. *BMJ* 2011;342:c5347.
- Bouri S, Shun-Shin MJ, Cole GD, *et al.* Meta-analysis of secure randomised controlled trials of β -blockade to prevent perioperative death in non-cardiac surgery. *Heart* 2014;100:456–64.
- Krumholz HM, Ross JS, Presler AH, *et al.* What have we learnt from Vioxx? *BMJ* 2007;334:120.
- Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. *Nature* 2005;435:737–8.
- Lüscher TF. The codex of science: honesty, precision, and truth—and its violations. *Eur Heart J* 2013;34:1018–23.
- Smith R. The COPE Report 2000. What is research misconduct? <http://publicationethics.org/files/u7141/COPE2000pdfcomplete.pdf> (accessed 26 Oct 2015).
- Eyding D, Lelgemann M, Grouven U, *et al.* Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 2010;341:c4737.
- Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ* 2012;344:d7202.
- Tramèr MR, Reynolds DJ, Moore RA, *et al.* Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635–40.
- Selph SS, Ginsburg AD, Chou R. Impact of contacting study authors to obtain additional data for systematic reviews: diagnostic accuracy studies for hepatic fibrosis. *Syst Rev* 2014;3:107.
- Huss A, Egger M, Hug K, *et al.* Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environ. Health Perspect* 2007;115:1–4.
- Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, *et al.* Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: a systematic review of systematic reviews. *PLoS Med* 2013;10:e1001578.
- Chan AW. Out of sight but not out of mind: how to search for unpublished clinical trial evidence. *BMJ* 2012;344:d8013.
- Egger M, Smith GD, Schneider M, *et al.* Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629.
- von Elm E, Poggia G, Walder B, *et al.* Different patterns of duplicate publication: an analysis of articles used in systematic reviews. *JAMA* 2004;291:974–80.
- Weingarten MA, Paul M, Leibovici L. Assessing ethics of trials in systematic reviews. *BMJ* 2004;328:1013–14.
- Vergnes JN, Marchal-Sixou C, Nabet C, *et al.* Ethics in systematic reviews. *J Med Ethics* 2010;36:771–4.
- Vlassov V, Groves T. The role of Cochrane review authors in exposing research and publication misconduct. *Cochrane Database Syst Rev* 2010;2011:ED000015. (accessed 26 Oct 2015).
- von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453–7.
- Tramèr MR. The Boldt debacle. *Eur J Anaesthesiol* 2011;28:393–5.
- Kelley GA, Kelley KS, Tran ZV. Retrieval of missing data for meta-analysis: apractical example. *Int J Technol Assess Health Care* 2004;20:296–9.
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 2008;5:e217; discussion e217.
- Turner EH, Matthews AM, Linardatos E, *et al.* Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252–60.
- Hopewell S, Loudon K, Clarke MJ, *et al.* Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;(1):MR000006.
- Dwan K, Altman DG, Cresswell L, *et al.* Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database Syst Rev* 2011;(1):MR000031.
- Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;(2):MR000005.
- Doshi P, Dickersin K, Healy D, *et al.* Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
- Wager E, Elia N. Why should clinical trials be registered? *Eur J Anaesthesiol* 2014;31:397–400.
- Cook DJ, Guyatt GH, Ryan G, *et al.* Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993;269:2749–53.
- Chan AW, Hróbjartsson A, Haahr MT, *et al.* Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–65.
- Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
- Mullan RJ, Flynn DN, Carlberg B, *et al.* Systematic reviewers commonly contact study authors but do so with limited rigor. *J Clin Epidemiol* 2009;62:138–42.
- Lundh A, Sisono S, Lexchin J, *et al.* Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2012;12:MR000033. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.MR000033.pub2/abstract> (accessed 26 Oct 2015).
- Melander H, Ahlqvist-Rastad J, Meijer G, *et al.* Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171–3.
- Vedula SS, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of gabapentin. *PLoS Med* 2013;10:e1001378.
- Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454–65.
- Moher D, Liberati A, Tetzlaff J, *et al.* The PRISMA Group. Preferred Reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.



39. MECIR reporting standards now finalized | Cochrane Community (beta). <http://community.cochrane.org/news/tags/authors/mecir-reporting-standards-now-finalized> (accessed 26 Oct 2015).
40. Marret E, Elia N, Dahl JB, *et al.* Susceptibility to fraud in systematic reviews: lessons from the Reuben case. *Anesthesiology* 2009;111:1279–89.
41. Habre C, Tramèr MR, Pöpping DM, *et al.* Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection. *BMJ* 2014;349:g5219.
42. Elia N, Wager E, Tramèr MR. Fate of articles that warranted retraction due to ethical concerns: a descriptive cross-sectional study. *PLoS ONE* 2014;9:e85846.