Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?

Laura McAuley, Ba' Pham, Peter Tugwell, David Moher

Summary

Background The inclusion of only a subset of all available evidence in a meta-analysis may introduce biases and threaten its validity; this is particularly likely if the subset of included studies differ from those not included, which may be the case for published and grey literature (unpublished studies, with limited distribution). We set out to examine whether exclusion of grey literature, compared with its inclusion in meta-analysis, provides different estimates of the effectiveness of interventions assessed in randomised trials.

Methods From a random sample of 135 meta-analyses, we identified and retrieved 33 publications that included both grey and published primary studies. The 33 publications contributed 41 separate meta-analyses from several disease areas. General characteristics of the meta-analyses and associated studies and outcome data at the trial level were collected. We explored the effects of the inclusion of grey literature on the quantitative results using logistic-regression analyses.

Findings 33% of the meta-analyses were found to include some form of grey literature. The grey literature, when included, accounts for between 4·5% and 75% of the studies in a meta-analysis. On average, published work, compared with grey literature, yielded significantly larger estimates of the intervention effect by 15% (ratio of odds ratios=1·15 [95% CI 1·04–1·28]). Excluding abstracts from the analysis further compounded the exaggeration (1·33 [1·10–1·60]).

Interpretation The exclusion of grey literature from meta-analyses can lead to exaggerated estimates of intervention effectiveness. In general, meta-analysts should attempt to identify, retrieve, and include all reports, grey and published, that meet predefined inclusion criteria.

Lancet 2000; **356:** 1228–31

Thomas C Chalmers Center for Systematic Reviews, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada (L McAuley MSc, B Pham MSc, D Moher MSc)**; Departments of Epidemiology and Community Medicine** (P Tugwell MD, D Moher)**, Medicine** (P Tugwell) **and Paediatrics** (D Moher)**, University of Ottawa**

Correspondence to: David Moher, Thomas C Chalmers Center for Systematic Reviews, Children's Hospital of Eastern Ontario Research Institute, Room R226, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada
(e-mail: dmoher@uottawa.ca)

Introduction

A meta-analysis is multifactorial. Decisions need to be made about how to handle various factors, such as language of publication, quality, and publication status, at the individual study level. Within the domain of publication status, a major factor to consider is the inclusion of grey literature (ie, studies that are unpublished, have limited distribution, and/or are not included in bibliographical retrieval system).[1]

The inclusion of grey literature in a meta-analysis may help to overcome some of the problems of publication bias, and provide a more complete and objective answer to the question under consideration. However, it has been reported that only 31% of published meta-analyses include grey literature.[2] This omission may be because the nature of grey literature makes its exclusion more convenient; it is difficult to retrieve, it is frequently incomplete, and its quality may be difficult to assess. We aim to provide empirical evidence about the impact of the exclusion of grey literature from meta-analyses on the estimate of intervention effectiveness.

Methods

Selection of meta-analyses

A sample of 135 meta-analyses were drawn randomly from an existing database of 455 meta-analyses of randomised clinical trials.[3,4] The database was established in 1996 through MEDLINE searches from 1966 to 1995 with a detailed search strategy assembled with the aid of an information specialist.

Eligibility criteria

To be eligible, a study has to be deemed a meta-analysis (included pooled analyses of the results of several independent primary studies), the associated studies had to be identifiable, and at least one item of grey literature (abstracts, unpublished studies, conference proceedings, graduate theses, book chapters, company reports, and applications) and one item of published work had to have been used in the generation of a summary statistic of the intervention effect. For reasons of feasibility, only meta-analyses that included binary outcomes and fewer than 100 randomised trials were considered.

Data abstraction

The following data were extracted via a structured form: the number of randomised trials; language of publication of these trials; year of publication of the meta-analysis and associated trials; number of patients; number and sources of grey literature; clinical area; outcome data. In meta-analyses that reported a positive outcome (such as survival) a complement outcome variable was computed (group–survivors=deaths).

Outcome data (number of unwanted events and total patients in the treatment and control groups) were extracted for all independent comparisons (non-overlapping randomised trials) from the published meta-analyses or from the original trials, when necessary, by one of us, and a subset was reviewed by another. Consensus between the two was achieved for any discrepancies before data entry.

	MA level (n=33)
Number of RCTs per MA (median [IQR])	10 (6–19)
Number of patients per MA (median [IQR])	1463 (1120–3163)
Clinical area (frequency)	
Gastrointestinal	24·2%
Cardiac	21·2%
Infection	12·1%
Reproduction	12·1%
Circulatory	9·1%
Number of sources of grey literature per MA (median [IQR])	1 (1–2)
Number of grey items per MA (median [IQR])	2 (1–3)
Year of publication of MA (median [IQR])	1993 (1991–1994)
Interventon comparisons	
Published MA with one comparison	27
Published MA with two independent comparisons	4
Published MA with three independent comparisons	2

MA=meta-analysis. RCT=randomised clinical trial.

Table 1: **General characteristics of meta-analyses**

The main outcome was defined as the one stated as such by the investigators, or if there was no such statement, the most clinically relevant was used (mortality would be selected over morbidity); if no outcome was clinically relevant, then the one contributing the most patients was used.[5]

Data analyses

We replicated the published meta-analyses as a quality-control measure. After completing the replication, each meta-analyses was repeated with all grey items removed, which allowed us to examine the effect of grey literature on estimates of the intervention effect. By the use of a logistic-regression model, the log odds of unwanted events experienced by each treatment group were related to trial, intervention, meta-analysis, and grey literature. Differences in the intervention effect across the meta-analyses were accounted for by an interaction effect between intervention and meta-analysis. An interaction term between grey literature and intervention was included to capture the potential that grey literature could modify estimates of intervention effectiveness. This effect was expressed as a ratio of intervention effect odds ratios (ie, ratio of odds ratios [ROR]) between grey and published literature.[6,7] With this modelling convention, an odds ratio less than 1·0 showed that the intervention was more effective than the control in preventing an unwanted event. Consequently, an ROR between grey and published literature greater than 1·0 indicated that, on average, estimates of intervention effectiveness from grey literature were smaller than their corresponding estimates from published work. The results from the unconditional models described above were verified by a conditional logistic-regression approach (not reported). We also completed a sensitivity analysis—we assessed the potential effect of abstracts as the source of grey literature, relative to published literature.

Standard residual diagnostics were used to assess the model's goodness of fit. ROR and its 95% CIs were derived from the fitted model. To assess the impact of grey literature on the variable "no intervention effect", Z scores were derived (ie, intervention effect size divided by its SE) from a meta-analysis with and without grey literature. The Z scores were then compared statistically by paired *t* test. The conditional logistic-regression analyses (data not reported) were done in LogXact, all other analyses in S-Plus 2000.

Results

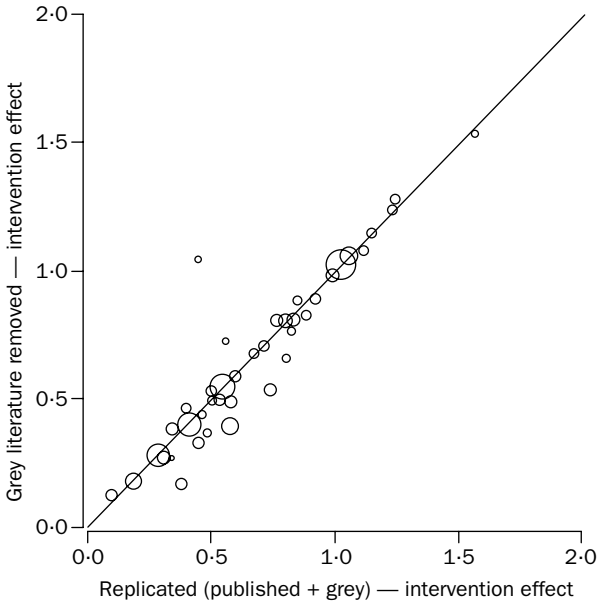
From the original 135 meta-analyses retrieved, we could not establish whether grey literature was included in 12 cases (11 poorly referenced, one not retrieved). An additional eight meta-analyses were excluded because they did not fit the definition of a meta-analysis. Of the remaining 115, 38 (33%) were found to contain at least one item of grey literature. Five of these meta-analyses failed to meet the inclusion criteria. For all subsequent investigations, we used the sample of 33 publications, contributing 41 separate and independent meta-analyses (table 1).

When included, grey literature accounted for between 4·5% and 75% (median=25% [interquartile range= 16·7–33·3%]) of the studies in a meta-analysis. For the sample as a whole, grey literature contributed 22% (102 of 467) of the studies and 10·7% (23 286 of 217 427) of the participants (table 2). Grey-literature studies tended to be smaller than published ones. None of the meta-

	RCT level		
	Overall	Grey	Published
Total number of RCTs	467	102	365
Total number of patients	217 427	23 286	194 141
Number of patients per RCT (median [IQR])	106 (55–223)	83·5 (48–190)	113 (59–228)
Number with null or negative results	157 (34·6%)	32 (32·0%)	125 (35·3%)
Year of publication* (median [IQR])	1988 (1985–1990)	1989 (1986–1990)	1988 (1984–1990)
Language of publication			
English	439 (94·0%)	99 (97·1%)	340 (93·2%)
French	11 (2·4%)	2 (2·0%)	9 (2·5%)
German	9 (2·0%)	0	9 (2·5%)
Spanish	4 (0·9%)	1 (1·0%)	3 (0·8%)
Italian	3 (0·7%)	0	3 (0·8%)
Japanese	1 (0·2%)	0	1 (0·3%)
Frequency of grey sources			
Abstracts & symposia	62%		
Unpublished	17%		
Book chapters	6%		
Reports	5%		
Pharmaceutical companies	3%		
In press	3%		
Letters	2%		
Theses	2%		

RCT=randomised clinical trial. *Publication indicates a prepared report for all grey items.

Table 2: **General characteristics of randomised trials included in the meta-analyses**



Pooled odds ratios with grey literature included plotted against corresponding odds ratios of same meta-analyses after removal of grey literature

Size of plotting circles is inversely proportional to variance of "grey literature removed" estimates.

analyses included more than three sources of grey literature. Most only included one source, usually abstracts, which accounted for 61% of the grey literature. In the sample as a whole, most of the items were in English (94·0%); this trend remained when grey and published studies were examined individually.

The replicated estimates of intervention effectiveness were very similar to published ones. Of the 41 meta-analyses, discrepancies of greater than 10% occurred only in one case. In this case the published report[8] shows a slightly higher worsening of suicidal ideation, although both the published and the replicated results are significant. In a meta-analysis on the use of β-blockers to prevent mortality in patients with cirrhosis and endoscopic varices,[9] the replication yielded a significant result in favour of treatment (OR=0·71 [95% CI 0·53–0·95]) when the published meta-analysis had reported a non-significant result (0·75 [0·57–1·06]).

In 14 (34%) of the 41 analyses removal of the grey literature changed the estimate of intervention effectiveness by 10% or more. In nine of these cases, removal resulted in the intervention effect moving away from unity. In three meta-analyses, the exclusion of the grey literature resulted in a change in the significance of the results, from non-significant to significant, in two cases. On average, across 39 meta-analyses (two were excluded because they contained non-independent trials), grey literature, compared with published, yielded a significantly larger odds-ratio estimate of treatment effect (15%, ROR 1·15 [1·04–1·28]), indicating that treatment is less effective for preventing an undesirable health outcome. The implication of this average effect on the individual meta-analysis is presented in the figure. When we limit the analysis to a comparison of abstracts and full publications, the estimate of intervention effectiveness (ROR 1·02 [0·91–1·14]) did not change. When abstracts were removed from this analysis (20 meta-analyses), this overestimate increased to 33% (1·33 [1·10–1·60]).

A trend towards more significant results after the removal of grey literature was found. When the Z score calculated with the grey literature excluded was compared with the Z score of the RORs (published and inclusive of grey literature) there was a significant decrease in Z scores (t=−7·257, p<0·001). In the sensitivity analysis, when abstracts were removed as a source of grey literature, the trend persisted (t=−6·422, p<0·001).

To illustrate the effect grey literature can have on the estimate of the effectiveness of an intervention in an individual meta-analysis, we examined the treatment of chronic venous insufficiency with hydroxyethylrutosides to reduce persistent leg pain.[10] The investigators reported relief in leg pain, but because we wanted all outcomes reported as negative events, we inferred lack of relief in leg pain from the presented data. This meta-analysis includes 13 randomised trials, of which ten were published in journals, two were internal reports, and one was in preparation for publication. Pooling all 13 studies, the reduction in persistent pain in those patients receiving treatment was 42% (OR=0·58 [95% CI 0·46–0·72]). When the three grey items are removed from the analysis, the reduction in pain relief increased to 61% (0·39 [0·30–0·52]), a 19% increase in pain relief.

Discussion

Our results suggest that the exclusion of grey literature from meta-analysis may result in an overestimate of an intervention effect by an average of 12%. In respect of quinine for nocturnal leg cramp, Hing and colleagues published a meta-analysis of four published studies.[11] These investigators then repeated their meta-analysis with three previously unidentified US Food and Drug Administration (FDA) documents. They report that the published trials, compared with the FDA documents, consistently reported larger estimates for the efficacy of quinine by about 50%. Similar evidence has been reported elsewhere (C MacLean, personal communication).

In the sensitivity analysis, when we removed abstracts from the sample, the overestimate increased to 38%. Abstracts compared with published literature had no impact on the point estimate of treatment effectiveness. This finding suggests that the effect of grey literature is altered by the inclusion of abstracts. There are strong arguments for not including abstracts with the grey literature—they may be catalogued on electronic databases (one of the 61 abstracts was identified through a search of PubMed); they are usually reported, at least in conference proceedings; they may be peer-reviewed; they may be more easily retrieved than other sources of grey literature. However, it takes, on average, 2–8 years for an abstract to be published as a full manuscript, and only about half ever appear as full papers.[12,13] Abstracts do not escape publication bias—ie, abstracts with positive findings tend to be accepted for presentation at conferences more frequently than those with negative or null findings.

The general understanding among meta-analysts is that grey literature should be sought and included in systematic reviews. However, 30% of editors surveyed would not publish a meta-analysis that included unpublished material,[2] even if it received favourable review. This negative view of grey literature may stem from the fact that it has not been peer reviewed. Although this perception may be well founded, we are not aware of any strong evidence, such as randomised trials, indicating the effectiveness of peer review. There is some evidence that grey literature is of lower quality than published literature.[15]

Some limitations of this study are actually limitations or deficiencies of the published meta-analyses. Poor referencing of included trials led to the exclusion of eleven meta-analyses from our sample because we could not find out which studies were used to generate the summary estimates. The inclusion of grey literature was incomplete in at least one of the included meta-analysis. The originators of this meta-analysis clearly state their inclusion criteria: "except for two unpublished references, which were obtained from the Diarrhoeal Disease Control Programme of the World Health Organization (WHO), only published reports were considered in an attempt to ensure the quality of the research".[16] Although this view is biased in terms of grey literature, it allows the reader to judge decisions. It is likely that there were similar conscious or unconscious omissions of grey literature in other meta-analyses in this sample. However, we could not judge how often this happened.

This work does not address the identification and retrieval of grey literature for those wishing to include it in future meta-analyses. It may be that the time, effort, and cost involved in identifying, locating, and retrieving the grey literature makes its inclusion prohibitive. This is an important area for future research. Efforts made by various groups, including the Cochrane Collaboration, through trial registries, negative trial journals, internet-based grey literature resources, and the policies adopted by several pharmaceutical companies to prospectively register trials and make these registers publicly available,[17] may make the identification easier. If investigators are expected to include grey literature, some guidance should be available on how to identify and retrieve it.

This work has implications for both meta-analysts and those who use them to help inform clinical and policy decisions. In general, the meta-analyst should attempt to ensure a comprehensive literature search to avoid the effects of selective publication. For those using meta-analyses to assist with clinical and policy decisions if grey literature is not included health-care decisions may be based on overly optimistic estimates of treatment effectiveness, which may have direct consequences for patient care.

Contributors

Laura McAuley completed the research as part of her MSc thesis and as such she helped develop the idea, collect the data, participated in all aspects of the data analyses and write-up of the paper. Ba' Pham supervised the data analyses and did some of them. Peter Tugwell was one of Laura's thesis supervisors and helped with the interpretation of the results, and David Moher helped to develop the idea, was one of Laura's thesis supervisors, helped to interpret the results, and participated in writing the paper.

Acknowledgments

We thank Jon Deeks for commenting on an earlier version of this paper, and Kay Dickersin, Ann McKibbon, and Jesse McGowan for their input into the definition of grey literature, and Robert Platt and Tong Guo for help running the conditional logistic-regression models.

References

1 Last JM. A dictionary of epidemiology, 3rd edn. Oxford: Oxford University Press, 1995.
2 Cook DJ, Guyatt GH, Ryan G, et al. Should unpublished data be included in meta-analyses? *JAMA* 1993; **269:** 2749–53.
3 Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998; **280:** 278–80.
4 Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352:** 609–13.
5 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; **272:** 122–24.
6 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273:** 408–12.
7 Khan KS, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril* 1996; **65:** 939–45.
8 Beasley CM Jr, Dorseif BE, Bosomworth JC, et al. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *BMJ* 1991; **303:** 685–92.
9 Pagliaro L, D'Amico G, Sorensen TI, et al. Prevention of first bleed in cirrhosis: a meta-analysis of randomized trials of nonsurgical treatment. *Ann Intern Med* 1992; **117:** 59–70.
10 Poynard T, Valterio C. Meta-analysis of hydroxyethylrutosides in the treatment of chronic venous insufficiency. *Vasa* 1994; **23:** 244–50.
11 Man-Son-Hing M, Wells G, Lau A. Quinine for nocturnal leg cramps: a meta-analysis including unpublished data. *J Gen Intern Med* 1998; **13:** 600–06.
12 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337:** 867–72.
13 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; **315:** 640–45.
14 Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998; **280:** 254–57.
15 McAuley LM. The influence of 'grey' literature on meta-analysis. Ottawa: University of Ottawa, 1999 (thesis).
16 Brown KH, Peerson JM, Fontaine O. Use of nonhuman milks in the dietary management of young children with acute diarrhea: a meta-analysis of clinical trials. *Pediatrics* 1994; **93:** 17–27.
17 Sykes R. Being a modern pharmaceutical company. *BMJ* 1998; **317:** 1172.