Epidemiology and Reporting Characteristics of Systematic Reviews

David Moher[1,2,3]*, Jennifer Tetzlaff[1], Andrea C. Tricco[1,4], Margaret Sampson[1], Douglas G. Altman[5]

1 Chalmers Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada, **2** Department of Paediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Canada, **3** Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada, **4** Institute of Population Health, University of Ottawa, Ottawa, Canada, **5** Centre for Statistics in Medicine, Oxford, United Kingdom

ABSTRACT

Background

Systematic reviews (SRs) have become increasingly popular to a wide range of stakeholders. We set out to capture a representative cross-sectional sample of published SRs and examine them in terms of a broad range of epidemiological, descriptive, and reporting characteristics, including emerging aspects not previously examined.

Methods and Findings

We searched Medline for SRs indexed during November 2004 and written in English. Citations were screened and those meeting our inclusion criteria were retained. Data were collected using a 51-item data collection form designed to assess the epidemiological and reporting details and the bias-related aspects of the reviews. The data were analyzed descriptively. In total 300 SRs were identified, suggesting a current annual publication rate of about 2,500, involving more than 33,700 separate studies including one-third of a million participants. The majority (272 [90.7%]) of SRs were reported in specialty journals. Most reviews (213 [71.0%]) were categorized as therapeutic, and included a median of 16 studies involving 1,112 participants. Funding sources were not reported in more than one-third (122 [40.7%]) of the reviews. Reviews typically searched a median of three electronic databases and two other sources, although only about two-thirds (208 [69.3%]) of them reported the years searched. Most (197/295 [66.8%]) reviews reported information about quality assessment, while few (68/294 [23.1%]) reported assessing for publication bias. A little over half (161/300 [53.7%]) of the SRs reported combining their results statistically, of which most (147/161 [91.3%]) assessed for consistency across studies. Few (53 [17.7%]) SRs reported being updates of previously completed reviews. No review had a registration number. Only half (150 [50.0%]) of the reviews used the term "systematic review" or "meta-analysis" in the title or abstract. There were large differences between Cochrane reviews and non-Cochrane reviews in the quality of reporting several characteristics.

Conclusions

SRs are now produced in large numbers, and our data suggest that the quality of their reporting is inconsistent. This situation might be improved if more widely agreed upon evidence-based reporting guidelines were endorsed and adhered to by authors and journals. These results substantiate the view that readers should not accept SRs uncritically.

The Editors' Summary of this article follows the references.

Introduction

Systematic reviews (SRs) have become increasingly popular in medicine. Clinicians read them as an efficient way of keeping up-to-date with their content area. Clinical practice guideline developers use them as a starting point for guideline development. Granting agencies require them as an evidence base for the need to conduct new research [1], and healthcare journals are moving in the same direction [2]. These different stakeholders should be encouraged to use SRs, particularly if they are of high quality and have taken steps to minimize bias.

The principal opportunity to estimate how a SR was designed and conducted is by examining its report. In one of the first evaluations of SRs, in 1987 Sacks and colleagues examined 83 SRs regarding their adequacy of reporting, using 23 characteristics covering six domains [3]. The authors noted that reporting was generally poor, with between 1 and 14 characteristics adequately reported (mean = 7.7; standard deviation = 2.7). For example, only six (7%) of the 83 reports mentioned use of a protocol to conduct the SR. A 1999 update [4] of a 1987 publication [5] found little improvement in the quality of reporting of SRs over time. Others have recently reported similar results, namely, that the quality of reporting of SRs is less than optimal [6,7].

Some possible limitations of these evaluations are that they did not set out to examine epidemiological aspects of SR reports, such as their publication prevalence and different types of SRs being published, nor did they examine a representative sample of published SRs. Typically the focus of such examinations is in specific health care areas [8–11] or type of SR [12,13]. Also, some of these evaluations are relatively old and thus could not address new and emerging issues relevant for SRs, such as the frequency with which they are updated. Our sampling strategy allowed us to comprehensively examine recent SRs across all application areas.

We set out to capture a cross-sectional sample of all recently published SRs and examine them in terms of a broad range of epidemiological and reporting characteristics, including emerging issues not previously examined.

Methods

Definition

There is no standard definition of an SR. We counted a report as an SR if the authors' stated objective was to summarize evidence from multiple studies and the article described explicit methods, regardless of the details provided.

Search

We believed that one month's publications would provide a few hundred eligible studies that would be enough to give a reliable summary of the literature. The Cochrane library is reissued quarterly. November 2004 was selected as it was the most recent month including new Cochrane reviews.

MEDLINE was searched to identify potential systematic reviews entered into Medline in November 2004. We used Montori's balanced five-term search strategy (specificity > sensitivity [14]), and supplemented it with his balanced three-term strategy (sensitivity > specificity), modified slightly: Ovid MEDLINE 1966 to Feb Week 1 2005 (searched Feb 18 2004), (1) 200411$.ed; (2) limit 1 to English (3) 2 and (cochrane database of systematic reviews.jn. or search.tw. or meta-analysis.pt. or medline.tw. or systematic review.tw. or ((meta-analysis.mp,pt. or review.pt. or search$.tw.) and methods.ab.))

In the case of the three-term strategy we required that the string "methods" appear in the abstract. We further limited the search to SRs reported in English due to the additional resources required to translate SRs published in other languages. The search results were uploaded into an Internet-based management program (Systematic Review System).

Screening

Two members of the research team independently screened the citations (title and abstract). One reviewer subsequently screened the full text articles of citations meeting our definition and those in which eligibility remained unclear. A second reviewer independently screened a 10% random sample. When it was clear that the intent of the authors was a literature review (e.g., authors identified the review as a brief overview with no specific review question), as opposed to an SR, articles were excluded. All other articles were included.

Data Collection and Analysis

Data collection was completed using a form that included 51 questions. Here we summarize a subset focusing on epidemiological and descriptive characteristics, including those with a potential for bias (Tables 1 and 2). We classified the Cochrane Database of Systematic Reviews, where Cochrane reviews are published, as a specialty journal. Reviews were classified according to their focus: therapeutic (e.g., treatment or prevention), diagnostic/prognostic (e.g., clinical prediction rules), epidemiology (e.g., incidence, prevalence), or other (e.g., education).

Two reviewers independently pilot-tested the form on a separate set of five articles. Each article was screened independently by one of three team members with a 5% sample screened in duplicate by two reviewers. Any uncertainties were discussed among the data extractors, and conflicts were resolved by coming to consensus.

All data analyses were performed using SPSS (version 13; http://www.spss.com). The analysis was descriptive. Data are summarized as frequency or median and interquartile range (IQR). We completed the one a priori subgroup analysis comparing Cochrane reviews to non-Cochrane reviews in terms of their reporting characteristics with a potential for bias.

Results

Search

Our search identified 1,046 records (Figure 1). Initial screening excluded 291 records, including two duplicates. The remaining 758 full-text articles were retrieved for additional scrutiny, of which 458 proved ineligible. Agreement was excellent between reviewers across the screening phases (kappa = 0.86). The majority of articles (78.2%) were excluded because they did not report any methods explicitly. We thus included 300 eligible reports for full data extraction.

Prevalence of Publishing SRs

We identified 300 SRs that were indexed in Medline in November 2004. Cochrane reviews contributed nearly half of these reviews (125 [41.7%]). Given that Cochrane reviews are

Table 1. Epidemiology of Systematic Reviews

Category	Characteristic	Number (%), of $n = 300$
Total number of journals		132
Number of journals publishing	One review	102 (77.3)
	Two reviews	21 (15.9)
	Three reviews	6 (4.5)
	Four reviews or more	3 (2.3)
Journal impact factor by review	0.0–5.0	106 (35.3)
	5.1–10.0	19 (6.3)
	10.1–15.0	1 (0.3)
	>15	4 (1.3)
	Not found	35 (11.7)
	N/A	135 (45.0)
Year of publication	2004	288 (96.0)
	2003	6 (2.0)
	2002	1 (0.3)
	2001	3 (1.0)
	2000	1 (0.3)
	1999	1 (0.3)
Number of authors	1	24 (8.0)
	2–3	125 (41.7)
	4–6	128 (42.7)
	≥7	23 (7.7)
Country of corresponding author	UK	76 (25.3)
	US	68 (22.7)
	Australia	31 (10.3)
	Canada	28 (9.3)
	The Netherlands	17 (5.7)
	Germany	10 (3.3)
	New Zealand	10 (3.3)
	Other (<10 reviews/country, 30 countries)	60 (20.0)
Focus of review[a]	Therapeutic[b]	213 (71.0)
	Diagnosis/Prognosis	23 (7.7)
	Epidemiology	38 (12.7)
	Other (e.g., education, review of quality of RCTs)	46 (15.3)
Common ICD-10	Mental and behavioural disorders	40 (13.3)
	Diseases of the circulatory system	33 (11.0)
	Malignant neoplasms	22 (7.3)
	Infections and parasitic diseases	20 (6.7)
	Diseases of the digestive system	20 (6.7)
	Pregnancy, childbirth, and the puerperium	21 (7.0)

[a]Does not equal 100%, as some reviews compared interventions in more than one category.
[b]Therapeutic refers to reports in which review focus was treatment or prevention.
N/A, not applicable.
doi:10.1371/journal.pmed.0040078.t001

published and indexed quarterly, this corresponds to an estimated annual frequency of approximately 2,500 publications, of which about 20% are Cochrane reviews. Although the SRs were all indexed in November 2004 they had been published between 1999 and 2004 (2004 = 96%). One review was published in the Cochrane Library and a paper-based journal; both are included here.

Using our definition of an SR we had difficulty in deciding on whether 30 (10%) of the included reviews were actually systematic. We completed a post-hoc sensitivity analysis with and without these 30 reviews. The results did not change in a

meaningful way. As such all of the analyses presented here include all 300 SRs.

Overall Epidemiology Characteristics

The 300 reviews were published in 132 journals, with most journals publishing only one SR indexed during the month although three journals each published four or more SRs (Table 1). Most reviews were published in journals with impact factors of five or less, although four SRs were published in journals with a high impact factor (>15). Nearly half of the SRs (135 [45%]), including all of the Cochrane SRs, did not have an impact factor. The reviews included a median of four authors although 24 (8%) of the reviews were single authored (Tables 1 and 2). One-quarter (76 [25.3%]) of the corresponding authors were from the United Kingdom with four countries (UK, US, Australia, and Canada) accounting for two-thirds of the SRs published during the month (Table 1). The majority of the reviews (213/300 [77%]) were classified as therapeutic—of which slightly more than half (125/213 [59%]) were Cochrane reviews—38/300 (13%) as epidemiology reviews, and 23/300 (8%) as diagnostic/prognostic reviews. Forty-six (15%) of the 300 reports were categorized as other, such as assessing the quality of immunosuppression trials in kidney transplantation [15].

Nearly half (142 [47.3%]) of the reviews reported examining the health effects of pharmacological interventions with the only a small percentage (24 [8.0%]) of reviews reported examining complementary and alternative medicine. The most common review classification (International Classification of Diseases 10 [ICD-10]) was mental and behavioural disorders (40 [13.3%]).

Descriptive Characteristics

Almost all of the SRs (272 [90.7%]) were published in specialty journals (Table 2). Few reviews (53 [17.7%]) reported being updates of previously completed reviews, although, of the therapeutic reviews, more than one-third of the Cochrane ones were updates (47/125 [37.6%]); almost no non-Cochrane reviews were reported as updates (2 [2.3%]). The SRs included a median of 16 studies involving 1,112 participants, although this number varied considerably by review category. Therapeutic reviews were the smallest and, of these, Cochrane reviews included fewer studies (median = 8) and participants (median = 769) compared to non-Cochrane reviews (median number of studies = 23; median number of participants = 1,137). Those reviews with a diagnostic/prognostic focus were the largest by far (Table 2). One hundred sixty-one reviews (53.7%) reported combining the results of their included studies statistically (i.e., meta-analytical synthesis). Most reviews (166/227 [73.1%]) reported on some aspect of harms. However, few reviews (61/254 [24%]) considered economic issues in their report.

Reporting Characteristics Related to Potential for Bias

Half (150 [50%]) of the review reports used the term "systematic review" or "meta-analysis" in the title or abstract with little variation across review category. However, of the therapeutic reviews, Cochrane reviews were less likely to use "systematic review" and/or "meta-analysis" in the title or abstract (49 [39.2%]) compared to non-Cochrane reviews (60 [68.2]%). Although close to half (139 [46.3%]) of the reviews reported working from a protocol, therapeutic reviews were more likely to report using a protocol than were other types

Table 2. Descriptive Characteristics of Included Systematic Reviews

Category	Characteristics	Overall[a], n = 300	Therapeutic, n = 213		Diagnosis/ Prognosis[b], n = 23	Epidemiology, n = 38	Other n = 46
			Cochrane, n = 125[c]	Non-Cochrane, n = 88			
Journal type	General, n (%)	28 (9.3)	0 (0.0)	16 (18.2)	4 (17.4)	4 (10.5)	8 (17.4)
	Specialty, n (%)	272 (90.7)	125 (100.0)	72 (81.8)	19 (82.7)	34 (89.5)	39 (84.8)
Number of authors, median (IQR)		4 (2–5)	3 (3–4)	3.5 (2.3–5)	4 (4–6)	3 (2–5)	4 (2–5)
Update of a previous review, n (%)		53 (17.7)	47 (37.6)	2 (2.3)	0 (0.0)	2 (5.3)	5 (10.9)
Types of interventions[d]	Pharmacological conventional, n (%)	142 (47.3)	84 (67.2)	47 (53.4)	3 (13.0)	7 (18.4)	13 (28.3)
	Nonpharmacological conventional, n (%)	113 (37.7)	41 (32.8)	43 (48.9)	14 (60.9)	6 (15.8)	18 (39.1)
	Complementary and alternative medicine, n (%)	24 (8.0)	11 (8.8)	11 (12.5)	0 (0.0)	1 (2.6)	1 (2.2)
	No intervention, n (%)	42 (14.0)	0 (0.0)	3 (3.4)	6 (26.1)	24 (63.2)	12 (26.1)
Number of included studies	n	289	125	84	20	36	42
	Median (IQR)	16 (7–30)	8 (4–16)	23 (12–38)	39 (21–67)	31 (16–50)	27 (12–54)
Number of participants in included studies	n	215	114	56	12	23	22
	Median (IQR)	1,112 (322–3,750)	769 (191–2,151)	1,137 (593–4,028)	14,523 (2,555–86,792)	2,189 (1,074–19,670)	1,644 (668–34,333)
Quantitative synthesis performed, n (%)		161 (53.7)	86 (68.8)	42 (47.7)	11 (47.8)	14 (36.8)	13 (28.3)
Harms considered[e], n (%)		166/227 (73.1)	105/122 (86.1)	44/77 (57.1)	7/13 (53.8)	4/9 (44.4)	17/22 (77.3)
Economics (i.e., costs) considered[e], n (%)		61/254 (24.0)	38/124 (30.6)	13/83 (15.7)	7/20 (35.0)	1/12 (8.3)	10/32 (31.3)

[a]Totals in rows do not equal "overall" as some reviews had multiple foci.
[b]Eighteen diagnostic focus, two prognosis focus, two both.
[c]This total does not account for all Cochrane reviews in the sample; four Cochrane reviews were classified in "Other"
[d]Does not equal 100%, as some reviews contained data in more than one category.
[e]Denominator of fractions indicate the number of reports where the variable concerned was considered relevant to the review.
doi:10.1371/journal.pmed.0040078.t002

of reviews (Table 3). Within the therapeutic review category almost every Cochrane review reported a protocol (122/125 [97.6%]), but only a small minority of non-Cochrane reviews did (10/88 [11.4%]).

The majority (211/293 [72.0%]) of reviews reported study design eligibility criteria, although this varied across and within review category. Of the therapeutic reviews almost all of Cochrane ones included randomized controlled trials
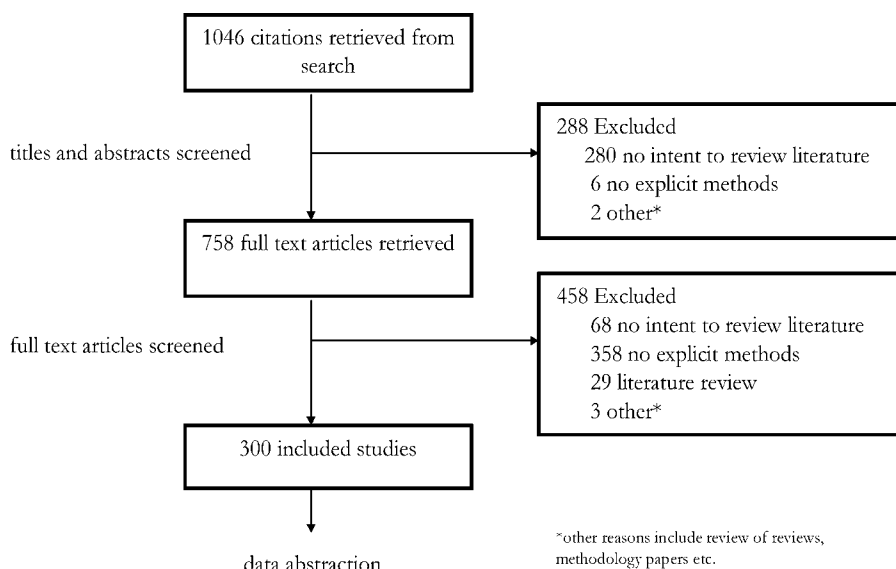
Figure 1. Flow of Citations through the Cross-Sectional Identification and Retrieval of Systematic Reviews Indexed in Medline during November 2004
doi:10.1371/journal.pmed.0040078.g001

(RCTs) only (123/125 [98.4%]), whereas slightly less than half of the non-Cochrane reviews had such restrictions (42/88 [47.7%]). Between 45% and 73% of the reviews in other categories did not specify any study design eligibility criteria (Table 3). About one-third of the reviews (109 [36.3%]) did not report whether they had any publication status eligibility criteria. Those reviews that did provide this information tended to include both published and unpublished documents (123 [41%]). Nearly half (134 [44.7%]) of the SRs did not report whether they had eligibility criteria based on language of reports. Of those that did provide this information, most (110 [36.7%]) included all languages (Table 3). Of the therapeutic reviews, only one Cochrane review (1/125 [0.8%]) reported limiting to reports of studies written in English, whereas about one third of non-Cochrane reviews reported this information (27/88 [30.7%]).

Reviews reported searching a median of three electronic databases and two other sources, although therapeutic reviews reported searching more sources than did other types of reviews (Table 3). For therapeutic reviews, Cochrane SRs typically searched twice as many databases compared to non-Cochrane SRs (median: four versus two, respectively). Approximately two-thirds (208 [69.3%]) of the reviews reported on the years covered for their electronic searches. There was little consistency concerning the electronic search information provided in SR reports. Of the therapeutic reviews, the majority of Cochrane reviews (98/125 [78.3%]) provided a full Boolean strategy, whereas only a minority of non-Cochrane reviews (16/88 [18.2%]) provided this information. Search strategies were reported in 10.9%–26.3% for the other review categories (Table 3). About half (145 [48.3%]) of the reviews reported on whether they considered duplicates as part of their overall approach to searching. For therapeutic reviews, twice as many Cochrane SRs (86/125 [68.8%]) considered duplicates compared to non-Cochrane reviews (30/88 [34.1%]).

Approximately half (143/280 [51.1%]) of the reviews reported at least one primary outcome, with reviews typically reporting a median of five outcomes, although therapeutic reviews reported nearly double this number (median = 9; see Table 3). For therapeutic reviews, the majority of Cochrane ones reported a primary outcome (97/125 [77.6%]), of which only a minority (14/97 [14%]) reported a statistically favourable result for this outcome whereas only about a third of non-Cochrane reviews reported a primary outcome (32/86 [37.2%]), of which half reported a statistically favourable result (16/32 [50%]). Two-thirds of the reviews (197/295 [66.8%]) reported information about quality assessment of their studies. For therapeutic reviews, all the Cochrane ones reported assessing the quality of included studies whereas only half of the non-Cochrane did (43/87 [49.4%]).

Most reviews (208 [69.3%]) reported on the flow of information throughout the review process, although there was little consistency as to how this flow was reported, with only a minority (20 [6.7%]) including the use of a QUOROM-like flow diagram (Table 3). Most reviews (250 [83.3%]) reported some information concerning the reasons for excluding studies, although, like the flow of information through the review, there was little consistency and completeness in how exclusions were reported (Table 3). Grey literature was reported to be included in 132 (44%) of SRs

with minimal differences across the review categories or within the therapeutic review category (Table 3).

Of those reviews reporting a quantitative synthesis, the vast majority (147/161 [91.3%]) reported assessing for consistency across the studies. However, less than one-quarter of the reviews (68/294 [23.1%]) reported assessing for publication bias. For therapeutic reviews, about twice as many Cochrane SRs (40/125 [32.0%]) reported assessing for publication bias compared to non-Cochrane reviews (16/87 [18.4%]). Funding sources were not reported in 122 (40.7%) of the reviews. A small number of reviews (7 [2.3%]) reported being funded by for-profit sources (Table 3). Of the therapeutic reviews that reported funding source information, nearly twice as many were Cochrane reviews (101/125 [80.8%]) compared to non-Cochrane reviews (44/88 [50%]). No SR in our sample reported a registration number.

Discussion

Our study shows that SRs are now being produced in large numbers, with about 2,500 new publications indexed annually on Medline, of which about one-fifth are Cochrane reviews. This value is an underestimate of the total number of new non-Cochrane SRs, as we examined only a single database and our searching might not have identified all SRs, and we restricted our study to English-language publications.

Our results indicate that SRs predominantly address questions about the effectiveness of interventions, and about half of them report combining their results statistically. Recent SRs include more studies and participants than previously reported [4–6], and while Cochrane reviews are smaller than other therapeutic reviews in our sample, their size corresponds to more recent data on Cochrane reviews [16]. Similarly, and like randomized trials [17], SRs are typically published in specialty journals.

There are some areas in which the quality of reporting has improved over time, such as the use of protocols [3] and quality assessment [18]. Similarly, very few reviews reported being funded by commercial sources, for which such funding has been associated with bias in the results of clinical trials [19].

It is too early to comment on the significance of some of our findings. Few reviews are reported as updates of previously completed reviews. The utility of reviews likely diminishes over time as they become outdated. Likewise, no review had a formal registration number. Our study also provided some disappointing results. For example, publication bias was considered or assessed in only a minority of reports despite much evidence for its existence and potential influence on the results of reviews [20,21].

However, for therapeutic reviews our comparison of Cochrane and non-Cochrane reviews provides the most discouraging results and suggests little improvement in the quality of reporting of non-Cochrane reviews over time [4,7]. The Cochrane Collaboration has a strict set of policies and guidance as to how SRs should be conducted and reported. For example, it is Cochrane policy not to include the words "systematic review" or "meta-analysis" in the review title. This policy explains why we observed fewer Cochrane reviews reporting these terms, compared to non-Cochrane ones. Similarly, Cochrane reviews, because they are reported in an electronic medium, are not encumbered by word length

Table 3. Descriptive and Reporting Characteristics with Potential for Bias

Category	Subcategory	Group	Overall[a], n = 300	Therapeutic, n = 213		Diagnosis/Prognosis[b], n = 23	Epidemiology, n = 38	Other, n = 46
				Cochrane, n = 125[c]	Non-Cochrane, n = 88			
	Use of terms "systematic review" or 'meta-analysis' in title or abstract		150 (50.0)	49 (39.2)	60 (68.2)	15 (65.2)	19 (50.0)	16 (34.8)
	Protocol mentioned		139 (46.3)	122 (97.6)	10 (11.4)	1 (4.3)	3 (7.9)	10 (21.7)
Eligibility criteria	Eligibility criteria based on study design[d,e]		211/293 (72.0)	125/125 (100.0)	55/88 (62.5)	6/22 (27.3)	14/36 (38.9)	23/42 (54.8)
		RCT	176 (60.1)	123 (98.4)	42 (47.7)	2 (9.1)	1 (2.8)	15 (35.7)
		Quasi-RCT or other controlled	72 (24.6)	51 (40.8)	15 (17.0)	2 (9.1)	0 (0.0)	6 (14.3)
		Cohort/case control	21 (7.2)	1 (0.8)	12 (13.6)	3 (13.6)	6 (16.7)	2 (4.8)
		Other observational	14 (4.8)	2 (1.6)	6 (6.8)	2 (9.1)	5 (13.9)	2 (4.8)
		Reviews/SR/MA	8 (2.7)	0 (0.0)	6 (6.8)	2 (9.1)	0 (0.0)	1 (2.4)
		Other or unclear	17 (5.8)	1 (0.8)	6 (6.8)	2 (9.1)	4 (11.1)	6 (14.3)
	Eligibility criteria based on publication status	Yes—published only	68 (22.7)	7 (5.6)	29 (33.0)	6 (26.1)	18 (47.4)	13 (28.3)
		Yes—unpublished only	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
		Yes—both published and unpublished	123 (41.0)	86 (68.8)	19 (21.6)	5 (21.7)	7 (18.4)	14 (30.4)
		Publication criteria not reported	109 (36.3)	32 (25.6)	40 (45.5)	12 (52.2)	13 (34.2)	19 (41.3)
	Eligibility criteria based on language of report	English only	49 (16.3)	1 (0.8)	27 (30.7)	7 (30.4)	10 (26.3)	12 (26.1)
		Mixed (English and LOE)	6 (2.0)	0 (0.0)	3 (3.4)	2 (8.7)	2 (5.3)	2 (4.3)
		Only LOE	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
		All languages considered	110 (36.7)	77 (61.6)	23 (26.1)	4 (17.4)	5 (13.2)	7 (15.2)
		Language criteria not reported	134 (44.7)	46 (36.8)	35 (39.8)	10 (43.5)	21 (55.3)	25 (54.3)
Search	Number of databases searched, median (IQR)		3 (2–5)	4 (3–6)	2 (2–4)	2 (2–4)	2 (1–3)	3 (1–5)
	Number of other sources searched, median (IQR)		2 (1–3)	3 (2–4)	2 (1–3)	2 (1–3)	1 (1–2)	1 (1–2)
	Years of coverage were reported	Yes	208 (69.3)	104 (83.2)	51 (58.0)	16 (69.6)	21 (55.3)	29 (63.0)
		Partially	49 (16.3)	16 (12.8)	18 (20.5)	2 (8.7)	5 (13.2)	9 (19.6)
		No search terms	37 (12.3)	1 (0.8)	19 (21.6)	3 (13.0)	10 (26.3)	5 (10.9)
	Search terms reported for one or more electronic databases[f]	Topics/free text	77 (25.7)	11 (8.8)	31 (35.2)	11 (47.8)	15 (39.5)	17 (37.0)
		Keywords/MeSH	51 (17.0)	14 (11.2)	19 (21.6)	4 (17.4)	7 (18.4)	10 (21.7)
		Full Boolean	127 (42.3)	98 (78.3)	16 (18.2)	4 (17.4)	4 (10.5)	13 (28.3)
		Readers are referred elsewhere for full search	5 (1.7)	1 (0.8)	3 (3.4)	1 (4.3)	0 (0.0)	0 (0.0)
Data abstraction	Duplicates considered		145 (48.3)	86 (68.8)	30 (34.1)	10 (43.5)	13 (34.2)	15 (32.6)
	One or more primary outcome(s) stated[e]		143/280 (51.1)	97/125 (77.6)	32/86 (37.2)	3/20 (15.0)	8/32 (25.0)	8/34 (23.5)
	Number of outcomes, median (IQR)		5 (2–9)	9 (6–14)	3 (2–7)	1 (1–2)	2 (1–5)	3 (1–7)
	Quality of primary studies assessed[e]		197/295 (66.8)	125/125 (100.0)	43/87 (49.4)	3/23 (13.0)	11/37 (29.7)	18/43 (41.9)
Results	Review flow reported[d]	No	92 (30.7)	17 (13.6)	38 (43.2)	9 (39.1)	20 (52.6)	16 (34.8)
		Partial information in table/text	99 (33.0)	60 (48.0)	17 (19.3)	5 (21.7)	9 (23.7)	10 (21.7)
		Yes, in table/text	106 (35.3)	47 (37.6)	31 (35.2)	9 (39.1)	9 (23.7)	20 (43.5)
		Yes, in QUOROM-like flow diagram	20 (6.7)	5 (4.0)	10 (11.4)	1 (4.3)	3 (7.9)	3 (6.5)
	Reasons for exclusion of studies reported[d]	No	50 (16.7)	3 (2.4)	22 (25.0)	7 (30.4)	12 (31.6)	12 (26.1)
		Partial information in table/text	119 (39.7)	49 (39.2)	33 (37.5)	7 (30.4)	19 (50.0)	16 (34.8)
		Yes, in table/text	126 (42.0)	69 (55.2)	32 (36.4)	9 (39.1)	7 (18.4)	18 (39.1)
		Yes, in QUOROM-like flow diagram	18 (6.0)	6 (4.8)	8 (9.1)	1 (4.3)	2 (5.3)	3 (6.5)
	Grey literature included	No	144 (48.0)	64 (51.2)	44 (50.0)	12 (52.2)	16 (42.1)	15 (32.6)
		Yes	132 (44.0)	59 (47.2)	35 (39.8)	11 (47.8)	17 (44.7)	22 (47.8)
		Can't tell/not reported	24 (8.0)	2 (1.6)	9 (10.2)	0 (0.0)	5 (13.2)	9 (19.6)
	Consistency (i.e. heterogeneity) investigated[e]		198/290 (68.3)	116/125 (92.8)	47/87 (54.0)	10/22 (45.4)	17/36 (47.2)	18/39 (46.2)
	Consistency investigated where synthesis performed[e]		147/161 (91.3)	86/86 (100.0)	35/42 (83.3)	8/11 (72.7)	11/14 (78.6)	12/13 (92.3)
	Publication bias assessed (or intent to assess)[e]		68/294 (23.1)	40/125 (32.0)	16/87 (18.4)	2/22 (9.0)	7/38 (18.4)	5/42 (11.9)
	Publication bias considered[e]		92/294 (31.3)	49/125 (39.2)	27/87 (31.0)	4/22 (18.2)	11/38 (28.9)	6/42 (14.3)

Table 3. Continued.

| Category | Subcategory | Group | Overall[a], n = 300 | Therapeutic, n = 213 | | Diagnosis/ Prognosis[b], n = 23 | Epidemiology, n = 38 | Other, n = 46 |
				Cochrane, n = 125[c]	Non-Cochrane, n = 88			
	Significance of primary outcome if one stated	Favourable significant	35/143 (24.5)	14/97 (14.4)	16/32 (50.0)	1/3 (33.3)	3/8 (37.5)	2/8 (25.0)
		Unfavourable significant	0/143 (0.0)	0/97 (0.0)	0/32 (0.0)	0/3 (0.0)	0/8 (0.0)	0/8 (0.0)
		Nonsignificant	38/143 (26.6)	32/97 (33.0)	5/32 (15.6)	0/3 (0.0)	1/8 (12.5)	1/8 (12.5)
		Unclear/mixed	70/143 (49.0)	51/97 (52.6)	11/32 (34.4)	2/3 (66.7)	4/8 (50.0)	5/8 (62.5)
Other	Funding sources reported	Nonprofit	144 (48.0)	81 (64.8)	32 (36.4)	6 (26.1)	10 (26.3)	19 (41.3)
		For-profit	7 (2.3)	1 (0.8)	4 (4.5)	0 (0.0)	1 (2.6)	1 (2.2)
		Mixed	19 (6.3)	9 (7.2)	5 (5.7)	2 (8.7)	2 (5.3)	2 (4.3)
		Author reported no funding	3 (1.0)	0 (0.0)	2 (2.3)	0 (0.0)	0 (0.0)	1 (2.2)
		Can't tell	5 (1.7)	4 (3.2)	1 (1.1)	0 (0.0)	0 (0.0)	0 (0.0)
	Funding sources not reported		122 (40.7)	24 (19.2)	44 (50.0)	15 (65.2)	25 (65.8)	23 (50.0)

All values given as n (%) except where indicated.
[a] Totals in rows do not equal "overall," as some reviews had multiple foci.
[b] Eighteen diagnostic focus, three prognosis focus, two both.
[c] This total does not account for all Cochrane reviews in the sample; four Cochrane reviews were classified as "Other."
[d] Denominator of fractions indicate the number of reports in which the variable concerned was considered relevant to the review.
[e] Does not equal 100%, as three reviews (two in epidemiological, one in other) did not search a major database.
RCT, randomized controlled trial; MA, meta-analyses; LOE, reports in languages other than English; MeSH, Medical Subject Headings; QUOROM, Quality of Reporting of Meta-analyses.
doi:10.1371/journal.pmed.0040078.t003

restrictions as is the case for most paper-based journals. Electronic publication, along with mechanisms to ensure adherence to Cochrane policies, allows authors to provide more complete details of how they conducted their SRs, possibly explaining why we observed far superior reporting standards of Cochrane reviews compared to non-Cochrane therapeutic ones. Journal editors and/or readers, particularly of specialty journals, might be less interested in these details and/or not have the resources to monitor adherence to any set of policies and reporting guidelines.

Many non-Cochrane reviews did not report key aspects of SR methodology, thus impairing confidence in their results and conclusions. For example, only 11% of the authors reported working from a protocol to complete their review, showing little improvement over time [3]. An examination of 47 Cochrane reviews, for which protocols almost always exist, revealed that 43 (91.5%) of them reported a major change such as the addition or deletion of outcomes between the protocol and the full publication [22].

Strong evidence of outcome reporting bias was recently reported within clinical trials [23,24]. Our results suggest that some aspect of selective outcome reporting bias might also exist within non-Cochrane reviews. Only about one-quarter of them reported a primary outcome, of which half report statistical significance in favour of this outcome (versus 14.4% for Cochrane reviews). This issue requires further investigation.

Results from SRs are most useful when they are up-to-date [25]. Few reviews were reported as updates and there were large differences between Cochrane reviews and non-Cochrane reviews. This difference might reflect differing or nonexistent policies and/or practices across funders, but also may reflect the reluctance of journals to publish updates that are substantially the same as previous publications. If these SRs are to retain their currency, updating them needs to be a much higher priority. This issue is likely to become increasingly important in coming years.

Clinical trial registration was called for twenty years ago [26]. In an attempt to minimize or avoid recent questionable behaviours such as hiding data [27], it is now becoming widely endorsed by granting agencies, such as the Canadian Institutes of Health Research [1], and editorial groups, such as the International Committee of Medical Journal Editors [28]. Similar behaviours may well affect SRs, although there are currently few data to inform this belief. Although protocols of Cochrane reviews are published, no review in our sample was registered in the usual sense, or had any type of registration number. Systematic review registration options exist [29,30] but are not well known. However, the reasons for registering SRs are likely different from that of registering clinical trials.

Providing reporting guidance is one way to improve the above-mentioned deficiencies. Real improvements have been seen in the reporting of RCTs [31] since the introduction of the CONSORT Statement and endorsement by journals [32,33] and editorial groups (for example, http://www.icmje. org). The QUOROM Statement, which provides reporting guidance for meta-analysis of RCTs, was developed ten years ago in the hope of having a similar impact on meta-analysis [34], and there are some indications that its use is associated with improved quality of reporting [35]. Its uptake was much slower than was the CONSORT statement, possibly reflecting

the belief that SRs were less frequently reported and not as important as RCTs. However, our study shows that currently about 2,500 such studies are published annually in English and indexed in Medline. QUOROM is currently being revised, and when published will provide journals and others with another opportunity to endorse its use as one way to improve the reporting of SRs. It is possible that the poor reporting of some SR categories reflects inadequate guidance available to authors. For example, we are unaware of any reporting guidance for SRs of prognostic studies.

Efforts directed toward improving reporting standards come at the end of the research process. Attempts to exert an influence much earlier in the research cycle might have a more profound and lasting effect. One option is to ensure that the evaluation process by granting agencies requires prospective applicants and grant peer reviewers to have some form of experience and expertise with SRs, for example by providing documentation of SR training. Additionally, journals might consider asking authors to include their SR protocols along with the completed review submissions.

Our cross-sectional analysis was limited in that we included just one month from a single database. It is unlikely, however, that sampling other months would have changed our results, as we verified that the month we chose was a typical entry month. While we included 300 reviews in our sample the majority of them were categorized as therapeutic. As such, our results concerning the other categories should be interpreted cautiously.

Our examination relied on what the authors reported. It is possible that the authors more completely conducted their SRs but omitted important details from their report, or the peer-review process resulted in the removal of key information we sought.

We have not provided here detailed information on several characteristics, such as the methods used to perform quantitative data synthesis or assess publication bias. More detailed information will be forthcoming in additional reports. Our intent here is to provide the reader with a broad overview of the reporting characteristics in recently published SRs.

Finally, we used a broad definition of what constitutes a SR. It is possible that a small proportion of our sample were intended to be literature reviews with very thorough descriptions of methods employed for the literature search. However, as such reviews would likely be used to the same extent as reports that are clearly defined as SRs or whose authors have quantitatively synthesized the information, we feel that they are subject to the same susceptibility to bias. Despite publishing SRs for about half a century, the National Library of Medicine does not yet index them as a publication type, at least in part because there is no agreed-upon definition of what constitutes a SR [36], although there are initiatives to help identify them [37]. A publication type for SRs on Medline is urgently needed.

Acknowledgments

References

1. Canadian Institutes of Health Research. Available at: http://www.cihr-irsc.gc.ca/e/documents/rct__reg__e.pdf. Accessed 24 January 2006.
2. Young C, Horton R (2005) Putting clinical trials into context. Lancet 366: 107.
3. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC (1987) Meta-analysis of randomized controlled trials. N Engl J Med 316: 450–455.
4. McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson F, et al. (1999) The medical review article revisited: Has the science improved? Ann Intern Med 131: 947–951.
5. Mulrow CD (1987) The medical review article: State of the science. Ann Intern Med 106: 485–488.
6. Jadad AR, McQuay HJ (1996) Meta-analyses to evaluate analgesic interventions: A systematic qualitative review of their methodology. J Clin Epidemiol 49: 235–243.
7. Shea B, Moher D, Graham I, Pham BA, Tugwell P (2002) A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 25: 116–129.
8. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, et al. (2000) Systematic reviews and meta-analyses on treatment of asthma: Critical evaluation. BMJ 320: 537–540.
9. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, et al. (2005) A systematic evaluation of the quality of meta-analyses in the critical care literature. Crit Care 9: R575–R582.
10. Bhandari M, Morrow F, Kulkarni AV, Tornetta P (2001) Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. J Bone Joint Surg Am 83A: 15–24.
11. Glenny AM, Esposito M, Coulthard P, Worthington HV (2003) The assessment of systematic reviews in dentistry. Eur J Oral Sci 111: 85–92.
12. Hayden JA, Cote P, Bombardier C (2006) Evaluation of the quality of prognosis studies in systematic reviews. Ann Inter Med 144: 427–437.
13. Whiting P, Rutjes AW, Dinnes J, Reitsma JB, Bossuyt PM, et al. (2005) A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. J Clin Epidemiol 58: 1–12.
14. Montori VM, Wilczynski NL, Morgan D, Haynes RB (2005) Optimal search strategies for retrieving systematic reviews from Medline analytical survey. BMJ 330: 68.
15. Fritsche L, Einecke G, Fleiner F, Dragun D, Neumayer HH, et al. (2004) Reports of large immunosuppression trials in kidney transplanation: Room for improvement. Am J Transplant 4: 738–743.
16. Mallett S, Clarke M. (2002) The typical Cochrane review: How many trials, how many participants. Int J Technol Assess Health Care 18: 820–823
17. Chan AW, Altman DG (2005) Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 365: 1159–1162.
18. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, et al. (2005) Assessment of methodological quality of primary studies by systematic reviews: Results of the metaquality cross sectional study. BMJ 330: 1053–1058.
19. Lexchin J, Bero LA, Djulbegovic B, Clark O (2003) Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. BMJ 326: 1167–1170.
20. Dickersin K (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication Bias in Meta-Analysis—Prevention, Assessment and Adjustments. West Sussex (UK): John Wiley & Sons. 356 p.
21. Sutton AJ (2005) Evidence concerning the consequences of publication and related biases. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication Bias in Meta-Analysis—Prevention, Assessment and Adjustments. West Sussex (UK): John Wiley & Sons. 356 p.
22. Silagy CA, Middleton P, Hopewell S (2002) Publishing protocols of systematic reviews: Comparing what was done to what was planned. JAMA 287: 2831–2834.
23. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG (2004) Emperical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. JAMA 291: 2457–2465.
24. Chan AW, Krleza-Jeric K, Schmid I, Altman DG (2004) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 171: 735–740.
25. Moher D, Tsertsvadze A (2006) Systematic reviews: When is an update an update? Lancet 367: 881–883.
26. Simes RJ (1986) Publication bias: The case for an international registry of clinical trials. J Clin Oncol 4: 1529–1541.
27. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, et al. (2004) Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. Lancet 363: 1341–1345.
28. De Angelis C, Drazan JM, Frizelle FA, Haug C, Hoey J, et al. (2004) International Committee Medical Journal Editors. Clinical trial registration: A statement from the International Committee of Medical Journal Editors. CMAJ 171: 606–607.

29. Centre for Reviews and Dissemination. Available at: http://www.york.ac.uk/inst/crd/listong.htm. Accessed 22 January 2006

30. Australia The Joanna Briggs Institute. Available at: http://www.joannabriggs.edu.au/pubs/systematic__reviews__prot.php. Accessed 22 January 2006.

31. Plint A, Moher D, Altman DG, Schulz KF, Morrison A (2006) Does the CONSORT checklist improve the quality of reports of randomized controlled trials? A systematic review. Med J Aust 185: 263–267.

32. Moher D, Schulz KF, Altman DGfor the CONSORT group (2001) The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. Ann Intern Med 134: 657–662

33. Rennie D (1997) How to report randomized controlled trials: The CONSORT statement. JAMA 276: 649.

34. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. (1999) Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. Lancet 354: 1896–1900.

35. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, et al. (2005) A systematic evaluation of the quality of meta-analyses in the critical care literature. Crit Care 9: R575–R582.

36. Montori VM, Wilczynski NL, Morgan D, Haynes RB (2005) Optimal search strategies for retrieving systematic reviews from Medline: Analytical survey. BMJ 330: 68.

37. Schulman J-LA, Nelson SJ (2005) Systematic reviews and medical vocabulary: Rapid response to Montori, Wilczynski, Morgan et al. BMJ eLetters 330: 68#96833. Available at: http://www.bmj.com/cgi/eletters/330/7482/68#96283. Accessed 25 February 2007.

Editors' Summary

Background. In health care it is important to assess all the evidence available about what causes a disease or the best way to prevent, diagnose, or treat it. Decisions should not be made simply on the basis of—for example—the latest or biggest research study, but after a full consideration of the findings from all the research of good quality that has so far been conducted on the issue in question. This approach is known as "evidence-based medicine" (EBM). A report that is based on a search for studies addressing a clearly defined question, a quality assessment of the studies found, and a synthesis of the research findings, is known as a systematic review (SR). Conducting an SR is itself regarded as a research project and the methods involved can be quite complex. In particular, as with other forms of research, it is important to do everything possible to reduce bias. The leading role in developing the SR concept and the methods that should be used has been played by an international network called the Cochrane Collaboration (see "Additional Information" below), which was launched in 1992. However, SRs are now becoming commonplace. Many articles published in journals and elsewhere are described as being systematic reviews.

Why Was This Study Done? Since systematic reviews are claimed to be the best source of evidence, it is important that they should be well conducted and that bias should not have influenced the conclusions drawn in the review. Just because the authors of a paper that discusses evidence on a particular topic claim that they have done their review "systematically," it does not guarantee that their methods have been sound and that their report is of good quality. However, if they have reported details of their methods, then it can help users of the review decide whether they are looking at a review with conclusions they can rely on. The authors of this *PLoS Medicine* article wanted to find out how many SRs are now being published, where they are being published, and what questions they are addressing. They also wanted to see how well the methods of SRs are being reported.

What Did the Researchers Do and Find? They picked one month and looked for all the SRs added to the main list of medical literature in that month. They found 300, on a range of topics and in a variety of medical journals. They estimate that about 20% of reviews appearing each year are published by the Cochrane Collaboration. They found many cases in which important aspects of the methods used were not reported. For example, about a third of the SRs did not report how (if at all) the quality of the studies found in the search had been assessed. An important assessment, which analyzes for "publication bias," was reported as having been done in only about a quarter of the cases. Most of the reporting failures were in the "non-Cochrane" reviews.

What Do These Findings Mean? The authors concluded that the standards of reporting of SRs vary widely and that readers should, therefore, not accept the conclusions of SRs uncritically. To improve this situation, they urge that guidelines be drawn up regarding how SRs are reported. The writers of SRs and also the journals that publish them should follow these guidelines.

Additional Information. Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0040078.

- An editorial discussing this research article and its relevance to medical publishing appears in the same issue of *PLoS Medicine*
- A good source of information on the evidence-based approach to medicine is the James Lind Library
- The Web site of the Cochrane Collaboration is a good source of information on systematic reviews. In particular there is a newcomers' guide and information for health care "consumers". From this Web site, it is also possible to see summaries of the SRs published by the Cochrane Collaboration (readers in some countries can also view the complete SRs free of charge)
- Information on the practice of evidence-based medicine is available from the US Agency for Healthcare Research and Quality and the Canadian Agency for Drugs and Technologies in Health