Evaluating the Quality of Systematic Reviews in the Emergency Medicine Literature

Karen D. Kelly, BScN, PhD*
Andrew Travers, MD, MSc*
Marlene Dorgan, BA, MLS‡
Linda Slater, BEd, MLIS‡
Brian H. Rowe, MD, MSc*

From the Division of Emergency Medicine, Faculty of Medicine and Dentistry, and John W. Scott Health Sciences Library,‡ University of Alberta, Edmonton, Alberta, Canada.*

Dr. Kelly is currently affiliated with the Department of Rural Health, University of Melbourne, Shepparton, Victoria, Australia.

Author contributions are provided at the end of this article.

Received for publication August 15, 2000. Revisions received January 23, 2001, and January 30, 2001. Accepted for publication February 27, 2001.

Presented at the Society for Academic Emergency Medicine annual meeting, Boston, MA, May 1999.

This study was supported by the Division of Emergency Medicine, University of Alberta (Dr. Rowe and Dr. Travers) in Edmonton, Alberta, Canada. Dr. Kelly has received funding from the Research Excellence Envelope in the Faculty of Medicine and Dentistry at the University of Alberta.

The authors declare no known conflicts of interest. Dr. Kelly, Dr. Travers, and Dr. Rowe are active members of the Airway Review Group of the Cochrane Collaboration. Dr. Rowe is an author of one of the systematic reviews included in this study. Dr. Rowe has received research funding in the past from the following industry sponsors: Astra Pharma, Glaxo Wellcome, Pharmacia and Upjohn, Trudell, and Janssen-Ortho, but is not a paid consultant to any of these companies.

Reprints not available from the authors.

Address for correspondence: *Brian H. Rowe, MD, MSc, Division of Emergency Medicine, 1G1.63 Walter McKenzie Health Centre, 8440-112 St. Edmonton, Alberta, Canada, T6G 2B7; 780-407-6707, fax 780-407-3314; E-mail brian.rowe@ualberta.ca.*

0196-0644/2001/$35.00 + 0
47/1/115881
doi:10.1067/mem.2001.115881

See editorial, p. 566.

Study objective: The objective of this study was to examine the scientific quality of systematic reviews published in 5 leading emergency medicine journals.

Methods: MEDLINE and EMBASE databases were electronically searched to identify published systematic reviews. Searches were only conducted in emergency medicine journals during the past 10 years; 4 of the journals were also hand searched. Potential reviews were assessed independently by 2 reviewers for inclusion. Data regarding methods were extracted from each review independently by 2 reviewers. All systematic reviews were retrieved and rated for quality by using the 10 questions from the overview quality assessment questionnaire.

Results: Twenty-nine reviews were identified from more than 100 citations. The overall scientific quality of the systematic reviews was low (mean score, 2.7; 95% confidence interval 2.1 to 3.2; maximum possible score, 7.0). Selection and publication biases were rarely addressed in this collection of reviews. For example, the search strategies were only identified in 9 (31%) reviews, whereas independent study selection (6 [21%]) and quality assessment of included studies (9 [31%]) were infrequently performed. Overall, the majority of reviews had extensive flaws, and only 3 (10%) had minimal flaws.

Conclusion: The results of the study indicate that many of the systematic reviews published in the emergency medicine literature contain major flaws; reviews with poor methodology may limit the validity of reported results. Further efforts should be made to improve the design, reporting, and publication of systematic reviews in emergency medicine.

[Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH. Evaluating the quality of systematic reviews in the emergency medicine literature. *Ann Emerg Med.* November 2001;38:518-526.]

INTRODUCTION

Over the past 10 years, systematic reviews have gained popularity in the medical field as a method of synthesizing vast amounts of research evidence. The term "systematic review" is an umbrella term used to denote reviews that perform comprehensive searches designed to reduce selection and publication bias and also to evaluate included study quality. Not all systematic reviews pool data in a quantitative fashion (a so-called meta-analysis), but many do. Those that do not provide a pooled estimate are referred to as nonquantitative systematic reviews.

Therapy systematic reviews attempt to combine the results of individual clinical trials for the purposes of resolving conflict in the literature, enlarging sample sizes to increase statistical power, and investigating variations in different subgroups.[1] Although a large number of randomized clinical trials have been published, they are often small and insufficiently powered to detect important treatment benefits.[2] Therefore, systematic reviews can identify pooled results that clarify the evidence from the published literature.[1] They are critical tools with outcomes that can serve to lessen the gap between evidence and clinical practice in medical care, especially in fields like emergency medicine, in which large multicenter trials are infrequently performed.[3]

However, there is heated debate about the value of systematic reviews. Many believe that they represent the highest level of evidence on which to base treatment recommendations at or above the level of large randomized controlled trials.[4] Others are more critical and suggest that large clinical trials are the criterion standard to evaluate and establish effectiveness.[5] Still others point to the variable quality of systematic reviews, an observation made frequently in the randomized controlled trial literature. One approach to resolving this debate is to find, produce, and disseminate scientifically valid systematic reviews.

The scientific quality of systematic reviews has been the source of considerable research in the past. Although some controversy still remains, most would agree that the scientific quality of a review is closely linked to reducing the biases inherent in this form of research. Several important sources of bias have been identified and studied: selection, publication, and pooling biases. If these influences are not carefully considered and minimized, a systematic review can lead to invalid results and conclusions.[5,6] Publication bias can occur when nonsignificant studies remain unpublished or when the authors of the systematic review do not perform a thorough literature search. Selection bias arises when selection of studies for inclusion in the review is influenced by factors such as the results, the authors, or some other factor. Systematic reviews have also been criticized for pooling data from a heterogeneous sample of studies (with varying case mixes, disease severities, or comorbidities). Avoiding errors in systematic reviews requires specific methodologic approaches, and assessing the validity of the results requires explicit reporting of the methods used to conduct the review.[7]

Although emergency medicine journals have not traditionally focused on this form of research, the frequency of publication is increasing. The objective of this study was to examine the scientific quality of systematic reviews published in the leading emergency medicine journals. To date, a quantitative evaluation of systematic reviews published in emergency medicine journals has not been reported, although comparisons between Cochrane reviews and articles published in paper-based journals have been completed.[8] Research of this type is necessary to assist clinicians in the continual provision of evidence-based medicine.

MATERIALS AND METHODS

An expert search strategy was adapted from previous research[9] and applied to both MEDLINE and EMBASE databases. The emergency medicine journals selected for searching were the *Journal of Emergency Medicine (JEM)*, *American Journal of Emergency Medicine (AJEM)*, *Annals of Emergency Medicine (Annals)*, *Journal of Accident and Emergency Medicine (JAEM)*, and *Academic Emergency Medicine (AEM)*. An electronic search was performed and supplemented with hand searching in 4 of these journals (*Annals*, *AJEM*, *JEM*, and *AEM*). The search strategy is outlined in Table 1.

Abstracts of potential reviews were examined independently by 2 reviewers to assess each for inclusion in the study. We included systematic reviews that either pooled (quantitative) or did not pool (nonquantitative) their data. Clear exclusions were general or narrative reviews of topic areas, case reports with appended literature reviews, duplicate publications of the review or its data, economic evaluations, and reviews of rare case reports.

The complete text of articles deemed suitable for inclusion was reviewed independently by 3 reviewers to determine validity by using the overview quality assessment questionnaire (OQAQ) developed by Oxman et al.[10] The index is composed of 10 questions, with the first 9 questions focusing on 5 different aspects of scientific quality of a systematic review (search strategy, selection strategies,

quality assessment, pooling, and results). These questions are answered as "yes," "no," or "partially/can't tell." Question 10 evaluates the overall scientific quality of the review on a 7-point scale, with 1 and 2 indicating extensive flaws, 2 to 4 indicating major flaws, 4 to 6 indicating minor flaws, and 6 and 7 indicating minimal flaws. None of the scores from questions are weighted. The OQAQ has been psychometrically tested and found to be both valid and reliable.[10,11] A copy of the tool can be found in the Appendix.

The index was applied to each review independently by 3 researchers and then adjudicated by consensus. Further information on quality measures, focusing on publication bias, selection bias, and bias from combining the study results, was also collected by using a standardized data collection form specifically designed for this study. These data were extracted independently by 2 reviewers, and consensus was used to arrive at the final data entry information.

Data for all cases were analyzed by using SPSS software, version 7.5 (SPSS, Chicago, IL). Categoric variables are reported as counts and percentages. Continuous variables are reported as means with 95% confidence intervals (CIs), and associations between these variables were assessed by using Spearman correlation coefficients. The interrater agreement among the reviewers was calculated by using an intraclass correlation coefficient, and then the scores for each item were adjudicated. The results presented are those achieved after adjudication. For ease of interpretation, the OQAQ scores were grouped to delete overlapping of scores as follows: 1 and 2 indicate extensive flaws, 3 and 4 indicate major flaws, 5 and 6 indicate minor flaws, and 7 indicates minimal flaws.

Table 1.
Recommended search strategies for searching published literature for systematic reviews.[9]

1	meta analysis.pt.
2	exp meta-analysis/
3	metaanal*.tw.
4	(meta adj anal*).tw.
5	(quantitative* review* or quantitativ* overview*).tw.
6	(systematic* review* or methodol* overview*).tw.
7	(methodol* review* or analytic$* overview*).tw.
8	(analytic* review* or analytic* overview*).tw.
9	or/1-8
10	(medline or embase or index medicus).ti,ab,sh.
11	(pooled or pooling).ti,ab,sh.
12	([combined or combining] adj [data or trials or studies or reviews])
13	or/10-12
14	13 and review.pt.
15	9 or 14

pt, Publication type; **tw**, text word; **ab**, abstract; **sh**, subject heading.
*Expanded.

RESULTS

The MEDLINE search (Table 1) identified 29 potential review articles published in the 5 journals between January 1988 and December 1998, of which 23 (79%) were eligible for the study. A similar EMBASE search identified 36 potential review articles within the same time period, of which 18 (50%) were eligible for the study. Of the eligible review articles identified in MEDLINE, 74% of them were also found in EMBASE, and of the eligible review articles identified in EMBASE, 95% of them were also found in MEDLINE.

In total, 24 systematic reviews that met our criteria were captured with electronic search methods. Hand searching of 4 of the same journals between January 1988 and December 1998 retrieved an additional 5 review articles. Twenty-nine articles were identified for inclusion in the study: 19 (66%) meta-analyses and 10 (34%) nonquantitative systematic reviews.[12-40] The reviews published in the major emergency medicine journals during this 11-year period are outlined in Table 2. The number of such reviews published by the 5 journals are as follows: 9 in *Annals*, 8 in *AEM*, 7 in *AJEM*, 3 in *JEM*, and 2 in *JAEM*. Of note, the first systematic review was published in 1990, and this publication type did not become widespread in the emergency medicine literature until 1995.

The results of the application of the OQAQ index are summarized in Tables 3 and 4. Table 3 describes the validity assessment of the first 9 questions of the OQAQ, and Table 4 outlines the results of the overall scientific quality of the review (ie, question 10 of the OQAQ). OQAQ questions 1 and 2 (Table 3) assess the search strategies used in the review article; the search methods were clearly reported in 55% of the reviews. Questions 3 and 4 of the OQAQ assess the selection of trials included in the review; the inclusion criteria for deciding which trials to include in the review were reported in more than half of the reviews (69%). Bias in the selection of studies was clearly avoided in 21% of the reviews. Questions 5 and 6 of the OQAQ focus on included study validity. The validity of the primary studies was assessed by using appropriate criteria in approximately one fourth of the reviews (24%), and the authors referenced a validated quality scoring system in 17% of the reviews. Pooling of the data was appraised in questions 7, 8, and 9 of the OQAQ, and the

results indicate that the methods used to combine the findings of the relevant studies were reported in approximately half of the reviews (48%). Similarly, the findings of the relevant studies were combined appropriately, relative to the primary question, in 38% of the reviews. Finally, the conclusions made by the authors were supported by the data in 52% of the reviews.

Question 10 of the OQAQ grades the overall scientific quality of the review. Fifty-six percent of reviews (Table 4) were rated as having extensive flaws (score of 1 or 2), 31% had major flaws (score of 3 or 4), and 13% had minor flaws (score of 5, 6, or 7). No review received the maximum rating of 7. The mean overall quality score after adjudication was 2.7 (95% CI 2.1 to 3.2; Table 5). The 3 reviewers had similar assessments of quality scores. No association between year of publication and the overall quality score of the review could be identified (r^2=0.02, P=.51).

Table 6 outlines the measures used to determine the quality of the systematic reviews in more detail. The quality measures listed (search strategies, inclusion criteria, eliminating selection bias, validity assessment, eliminating validity bias, and pooling data) are characterized by using detailed description to further explain how the reviews were conducted. These quality measures are linked to the specific question of the OQAQ index, as indicated in Table 6.

DISCUSSION

This is the first study to evaluate the quality of systematic reviews in the emergency medicine literature. The results of this study suggest that searching for systematic reviews

Table 2.
Systematic reviews published in 5 emergency medicine journals during the past 10 years.

Reference	Publication Year	Topic	Type	Journal	Search Method*
Eisenberg et al[16]	1990	Therapy	MA	*Annals*	Hand
Hebert et al[24]	1991	Therapy	NQSR	*JEM*	MED/EM/hand
Campbell and Chyka[29]	1992	Therapy	MA	*AJEM*	Hand
Howell et al[32]	1992	Therapy	MA	*AJEM*	Hand
Hollander and Hoffman[17]	1992	Prevention	NQSR	*JEM*	MED/EM/hand
Rowe et al[22]	1992	Therapy	MA	*AJEM*	MED/EM/hand
Baraff et al[31]	1993	Therapy	MA	*Annals*	MED/EM/hand
Becker et al[25]	1993	Therapy	NQSR	*Annals*	MED/hand
Ward[30]	1994	Therapy	NQSR	*AEM*	MED/hand
Babbs[33]	1994	Therapy	NQSR	*AEM*	MED/hand
Cummings[14]	1994	Therapy	MA	*Annals*	MED/EM/hand
Morgan and Emerman[19]	1995	Therapy	NQSR	*AJEM*	MED/EM/hand
Auble et al[20]	1995	Therapy	MA	*Annals*	MED/EM/hand
Cummings and Beccaro[15]	1995	Therapy	MA	*AJEM*	MED/EM/hand
Orr et al[35]	1995	Diagnostic	MA	*AEM*	MED/hand
Osmond and Klassen[23]	1995	Therapy	MA	*AEM*	MED/hand
Watts[18]	1995	Therapy	MA	*Annals*	MED/EM/hand
Bonadio[13]	1996	Therapy	MA	*JEM*	EM/hand
Jabbour et al[28]	1996	Therapy	MA	*Annals*	MED/EM/hand
Maenza et al[12]	1996	Diagnostic	MA	*AJEM*	MED/EM/hand
Nichol et al[21]	1996	Therapy	MA	*Annals*	MED/EM/hand
Pearl and Todd[40]	1996	Diagnostic	NQSR	*Annals*	Hand
Bulloch et al[36]	1997	Therapy	MA	*AEM*	MED/EM/hand
Kinnane et al[26]	1997	Prevention	NQSR	*AEM*	MED/EM
Nee[34]	1997	Therapy	NQSR	*JAEM*	MED/hand
Walker and Slovis[27]	1997	Therapy	NQSR	*AEM*	Hand
Markert et al[38]	1998	Diagnosis	MA	*AJEM*	MED/EM/hand
Rothrock et al[39]	1998	Therapy	MA	*AEM*	MED/EM/hand
Leydon et al[37]	1998	Economic	NQSR	*JAEM*	MED/EM/hand

MA, Meta-analysis; **Hand**, hand search; **NQSR**, nonquantitative systematic review; **MED**, MEDLINE; **EM**, EMBASE.
*How the article was identified in the medical literature by the research team.

Table 3.
Validity assessment with the OQAQ (findings after adjudication).

Index	No	Partially/ Can't Tell	Yes
Q1. Search methods stated (%)	6 (21)	7 (24)	16 (55)
Q2. Search comprehensive (%)	14 (48)	10 (35)	5 (17)
Q3. Inclusion criteria reported (%)	3 (10)	6 (21)	20 (69)
Q4. Selection bias avoided (%)	9 (31)	14 (48)	6 (21)
Q5. Validity criteria reported (%)	21 (72)	3 (10)	5 (17)
Q6. Validity assessed appropriately (%)	13 (45)	9 (31)	7 (24)
Q7. Combining methods reported (%)	15 (52)	0	14 (48)
Q8. Finding combined appropriately (%)	9 (31)	9 (31)	11 (38)
Q9. Conclusions supported by data (%)	4 (14)	10 (35)	15 (52)

Table 4.
Overall scientific quality: Question 10 of the OQAQ (findings after adjudication).

Overall Quality Score	No. (%)
1	6 (21)
2	10 (35)
3	8 (28)
4	1 (3)
5	1 (3)
6	3 (10)
7	0
Total	29 (100)

in the emergency medicine literature can be a difficult task. For example, multiple search strategies are required, and computerized searches, which are thought to be sensitive, miss many systematic reviews. Practically, this creates problems for clinicians attempting to identify high-quality systematic reviews for use in teaching and practice. More importantly, searches with less specific terms will identify significant noise with many other review types, such as narrative reviews, state-of-the-art reviews, evidence-based medicine reviews, and other reviews.

When found, these reviews predominantly involve therapeutic issues and have been increasing in number since the mid-1990s. However, many of the systematic reviews published in emergency medicine journals over the past 10 years exhibit major methodologic flaws. For example, most do not include a comprehensive search of the published literature; almost all avoid unpublished and foreign-language literature. Many reviews did not address selection bias, and most do not assess the quality of studies included in the reviews. Although pooling is often attempted, few authors address the issue of heterogeneity. Overall, when assessed by using the OQAQ, the mean review scores are consistently low (2.7 out of a possible maximum score of 7.0). These results indicate that reviews published in the mainstream emergency medicine journals may contain biases, which, if not identified and accounted for, could lead to the incorporation of invalid findings into clinical practice.

Bias arising from publication has an important influence on the results of systematic reviews and is more difficult to detect and adjust for than other forms of bias. Thus, when searching for studies to include in a systematic review, it is critical that the literature search strategy is explicitly described, comprehensive, and inclusive of both unpublished and published research.[5] The description of the search strategy provides some assurance that the authors have conducted a comprehensive, detailed, and exhaustive search for literature relevant to the study question.

Although the term "comprehensive search" is a relative one, evidence exists to suggest that it should include computer searching of more than one database supplemented by reviewing bibliographies of relevant articles and contacting primary authors of relevant articles.[9] Simple MEDLINE computer searching has been shown to result in inadequate sensitivity because a simple MEDLINE search may yield less than 50% of all potentially relevant articles.[41] Although approximately two thirds of the reviews in emergency medicine journals searched the MEDLINE database, other databases were rarely searched, and attempts to identify unpublished literature were even less frequently completed. Finally, hand searching of journals, a method often used to identify additional trials,

Table 5.
Quality score of OQAQ by 3 independent reviewers.

Reviewer	Quality Score, Mean (95% CI)
Reviewer 1*	2.3 (1.9–2.7)
Reviewer 2*	3.0 (2.4–3.6)
Reviewer 3*	2.9 (2.3–3.6)
Pooled score	2.7 (2.2–3.3)
Adjudicated score	2.7 (2.2–3.2)

*These scores are from before adjudication.

Table 6.
Items used to assess scientific quality of systematic reviews in emergency medicine.

Quality Measures	Items	No. (%)
Search strategies (Q1 and Q2)*	MEDLINE	23 (79)
	Bibliographies	18 (62)
	Other computer sources	8 (28)
	Unpublished literature	4 (14)
	Contact with authors	4 (14)
	Hand searching	3 (10)
	Foreign language	2 (7)
	EMBASE	1 (3)
	Other	6 (21)
Inclusion criteria (Q3)	Inclusion	7 (24)
	Relevance	3 (10)
Eliminating selection bias (Q4)	Multiple assessors	7 (24)
	Assessors blinded	6 (21)
	Adjudication	6 (21)
	Measurement reported	4 (14)
	Papers blinded	0
	Other	7 (24)
Validity assessment (Q5)	Study quality reported	9 (31)
	Score used in sensitivity analyses	6 (23)
	Score used	5 (17)
	Reference provided	3 (10)
	Other	3 (10)
Eliminating validity bias (Q6)	Multiple assessors	4 (14)
	Assessors blinded	4 (14)
	Adjudication	4 (14)
	Measurement reported	2 (7)
Pooling data (Q7)	Pooling performed	18 (62)
	CI reported	12 (67)†
	Heterogeneity calculated and reported	8 (44)†
	Accounted for heterogeneity	8 (44)†

*Search strategies support questions 1 and 2 of the OQAQ.
†Denominators include only those reviews with pooled data.

was rarely used. Therefore, it is likely that the systematic reviews published in the emergency medicine literature suffer from publication bias.

Furthermore, it is necessary to explicitly state the criteria for including or excluding studies in a systematic review so that there is a clear understanding of what is being compared. The majority of reviews evaluated in this study did not describe the process by which articles were selected for relevance and inclusion in an unbiased manner. A 2-step process for identifying included articles is commonly used to eliminate selection bias. Often, the first step is to review the title, abstract, and medical subject headings terms to identify possibly relevant articles for retrieval. Each article is often then reviewed independently by at least 2 reviewers to determine whether it meets the inclusion criteria, and any disagreement is resolved. Because only a small percentage of authors reported a 2-step process for including studies in their reviews, it is possible that selection bias influenced the reported results.

The confidence in results generated by systematic reviews depends to a large extent on the quality of the studies included in the review. By including studies of low quality in systematic reviews, the results become more heterogeneous, and the pooled estimate of effect increases.[41] Therefore, the inclusion of trials of lower quality is more likely to produce positive results.[42] Ideally, the quality of the studies should be evaluated independently by at least 2 reviewers, and a calculation of interrater reliability is often reported. There are many scales, tools, and scoring systems currently used to assess the quality of clinical trials, yet only a few have been shown to be reliable and valid.[43] Clearly, debate exists about which approach to use when assessing trial quality. However, it appears more important that this process is actually performed and reported than that a specific method be used. Only 25% of the reviews published in emergency medicine journals assessed and reported the validity of the included trials.

Pooling of study results can produce serious errors in the estimation of effect size.[44] When combining the results of individual studies, the data should be evaluated for heterogeneity before combining. If significant heterogeneity exists, the pooled result should be viewed with caution, and reasons for the identified variability should be explored. More than 30% of the systematic reviews assessed in this study provided insufficient information to assess whether the findings were combined appropriately; however, when this information was available, the authors appeared to pool heterogeneous data in more than 25% of the reviews.

There are several limitations of this study. First, although all systematic reviews published in these journals were included in the study, this constituted a relatively small number of studies designated as reviews. Thus, this evaluation is confined to those articles that met strict criteria for inclusion. Many other articles in emergency medicine journals are designated as reviews and have no resemblance to systematic reviews. Therefore, this evaluation may overestimate the quality of reviews in the emergency medicine literature.

Second, although an extensive search for emergency medicine reviews was performed, we have not obtained all reviews published by emergency physicians in any journal. Clearly, emergency researchers publish systematic reviews in other venues, such as the Cochrane Library and non–emergency medicine journals. However, we believe that by searching for articles from 5 of the most widely read emergency medicine journals, we have reviewed those systematic reviews that would be most likely to influence the general emergency medicine community. Because of a lack of local access, *JAEM* was not hand searched; however, given the small number of systematic review articles identified in the electronic search of this journal, we do not believe this affects the results in any meaningful way.

Third, the methods used to evaluate the systematic reviews could be criticized. There are no formal criteria for rating reviews; however, we used a validated scoring system (OQAQ), used additional methods, and had at least 2 reviewers conduct these reviews independently. Furthermore, no authors reviewed their own research.[22] Finally, we measured agreement and found generally good agreement on most of the scores between reviewers. Therefore, we believe these scores represent a valid assessment of the current emergency medicine systematic review literature.

Fourth, the accuracy of the scoring suffers from the fact that the authors did not describe the methods used in sufficient detail. Although one solution could be to write to authors to clarify the methods, as is done in some systematic reviews, it seemed inappropriate given that readers are forced to make a decision on quality without the benefit of discussions with the systematic reviews' authors. We believe this was a more appropriate approach and justifies the comments made in the conclusion section.

Finally, although this study is the first to look at this type of data in the emergency medicine literature, comparison data from other disciplines are available. Previous reviews of systematic reviews in other fields have revealed similar methodologic concerns.[9,10,43] As in other set-

tings, there were several examples of high-quality methodologic research within the emergency medicine systematic review publications. Although none of the studies was rated as a 7 out of 7, there were some systematic reviews that conducted comprehensive searches, protected against selection bias, evaluated quality, and pooled studies appropriately. However, by using the OQAQ, this study revealed an overall low quality of systematic reviews, with more than 50% of the reviews exhibiting extensive flaws. Moreover, it does not appear that the quality of reviews has improved in more recent years. Therefore, improvements are required on a relatively urgent basis to change these statistics.

Notwithstanding the previously mentioned concerns, to our knowledge, this is the first study of the quality of published emergency medicine systematic reviews. The study demonstrated the difficulty experienced by many clinicians when searching for systematic reviews with sophisticated computer search methods. Once emergency medicine reviews were identified, in many instances, the methods sections were insufficiently detailed and therefore difficult to evaluate. The overall evaluation of these systematic reviews suggests that many systematic reviews in the emergency medicine literature contain major flaws.

In view of these results, we believe that specific changes must be made at multiple levels by publishers, authors, and readers of systematic reviews. First, journals should focus on accepting high-quality systematic reviews and on ensuring that the Methods sections outline the methods in a clearer manner.[45] Second, outlines for publishing systematic reviews need to be provided to authors, and it is necessary for systematic reviews to be evaluated on the basis of proven criteria. It is encouraging that some emergency medicine journals have recently engaged in discussions about the value of systematic reviews.[46,47] From this, we would anticipate editors providing guidelines for systematic review publication. However, this tactic will only be successful when editors permit appropriate depth in the description of the methods in the final article.

Authors also need to pay particular attention to methods used in systematic reviews, preferably before embarking on this research activity. The authors should address the potential biases of these methods in their discussion to provide the most valid estimation of effect.

Finally, readers need to become more familiar with critically appraising systematic reviews and developing a healthy skepticism before incorporating the results into practice. There are several excellent review-of-review articles available that can assist clinicians as guides to learning more about this area of medical publication.[2,7,48]

Emergency physicians require a firm understanding of systematic reviews to benefit from this potentially powerful research tool, and we encourage more discussion and integration of high-quality systematic reviews into emergency medicine practice.

Author contributions: KDK and BHR conceived the study and designed the protocol. MD and LS contributed to the search criteria and the quality scoring of any searches described in separate meta-analysis. KDK, AT, and BHR completed the independent OQAQ scoring; KDK and BHR independently completed data extraction. KDK supervised the entry and analysis of the data. KDK and BHR drafted the manuscript, and all authors contributed substantially to its revision. KDK and BHR take responsibility for the paper as a whole.

We thank Mrs. Diane Milette for her invaluable assistance in preparing this article for publication.

REFERENCES

1. Henry DA, Wilson A. Principles behind practice: meta-analysis. Part 1: an assessment of its aims, validity and reliability. *Med J Aust.* 1992;156:31-38.

2. Olson CM. Understanding and evaluating a meta-analysis. *Acad Emerg Med.* 1994;1:392-398.

3. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med.* 1994;120:667-676.

4. Guyatt GH, Sackett DL, Sinclair JC, et al. User's guide to the medical literature IX: a method of grading health care recommendations. *JAMA.* 1995;274:1800-1804.

5. Flather MD, Farkouh ME, Pogue JM, et al. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials.* 1997;18:568-579.

6. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA.* 1990;263:1392-1395.

7. Oxman AD, Cook DJ, Guyatt GH, et al. Users' guides to the medical literature VI. How to use an overview. *JAMA.* 1994;272:1367-1371.

8. Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses. *JAMA.* 1998;280:278-280.

9. Hunt DL, McKibbon KA. Locating and appraising systemic reviews. *Ann Intern Med.* 1997;126:532-538.

10. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol.* 1991;44:91-98.

11. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol.* 1991;44:1271-1278.

12. Maenza RL, Seaberg D, D'Amico F. A meta-analysis of blunt cardiac trauma: ending myocardial confusion. *Am J Emerg Med.* 1996;14:237-241.

13. Bonadio WA. Adjunctive dexamethasone therapy for pediatric bacterial meningitis. *J Emerg Med.* 1996;14:165-172.

14. Cummings P. Antibiotics to prevent infection in patients with dog bite wounds: a meta-analysis of randomized trials. *Ann Emerg Med.* 1994;23:535-540.

15. Cummings P, Beccaro MAD. Antibiotics to prevent infection of simple wounds: a meta-analysis of randomized studies. *Am J Emerg Med.* 1995;13:396-400.

16. Eisenberg MS, Horwood BT, Cummins RO, et al. Cardiac arrest and resuscitation: a tale of 29 cities. *Ann Emerg Med.* 1990;19:179-186.

17. Hollander JE, Hoffman RS. Cocaine-induced myocardial infarction: an analysis and review of the literature. *J Emerg Med.* 1992;10:169-177.

18. Watts DD. Defibrillation by basic emergency medical technicians: effect on survival. *Ann Emerg Med.* 1995;26:635-639.

19. Morgan B, Emerman CL. Effect of age on myocardial infarction and thrombolysis. *Am J Emerg Med.* 1995;13:196-198.

20. Auble TE, Menegazzi JJ, Paris PM. Effect of out-of-hospital defibrillation by basic life support providers on cardiac arrest mortality: a metaanalysis. *Ann Emerg Med.* 1995;25:642-648.

21. Nichol G, Detsky AS, Stiell IG, et al. Effectiveness of emergency medical services for victims of out-of-hospital cardiac arrest: a metaanalysis. *Ann Emerg Med.* 1996;27:700-710.

22. Rowe BH, Keller JL, Oxman AD. Steroid use in the emergency department treatment of asthma exacerbations: a meta-analysis. *Am J Emerg Med.* 1992;10:301-310.

23. Osmond MH, Klassen TP. Efficacy of ipratropium bromide in acute childhood asthma: a meta-analysis. *Acad Emerg Med.* 1995;2:651-656.

24. Hebert P, Weitzman BN, Stiell IG, et al. Epinephrine in cardiopulmonary resuscitation. *J Emerg Med.* 1991;9:487-495.

25. Becker LB, Smith DW, Rhodes KV. Incidence of cardiac arrest: a neglected factor in evaluating survival rates. *Ann Emerg Med.* 1993;22:86-91.

26. Kinnane JM, Garrison HG, Coben JH, et al. Injury prevention: is there a role for out-of-hospital emergency medical services? *Acad Emerg Med.* 1997;4:306-312.

27. Walker LA, Slovis CM. Lidocaine in the treatment of status epilepticus. *Acad Emerg Med.* 1997;4:918-922.

28. Jabbour M, Osmond MH, Klassen TP. Life support courses: are they effective? *Ann Emerg Med.* 1996;28:690-698.

29. Campbell JW, Chyka PA. Physicochemical characteristics of drugs and response to repeat-dose activated charcoal. *Am J Emerg Med.* 1992;10:208-210.

30. Ward KR. Possible reasons for the variability of human responses to IAC-CPR. *Acad Emerg Med.* 1994;1:482-489.

31. Baraff LJ, Bass JW, Fleisher GR, et al. Practice guideline for the management of infants and children 0 to 36 months of age with fever without source. *Ann Emerg Med.* 1993;22:1198-1210.

32. Howell JM, Dalsey WC, Hartsell FW, et al. Steroids for the treatment of corrosive esophageal injury: a statistical analysis of past studies. *Am J Emerg Med.* 1992;10:421-425.

33. Babbs CF. The evolution of abdominal compression in cardiopulmonary resuscitation. *Acad Emerg Med.* 1994;1:469-477.

34. Nee PA. Thrombolysis after acute myocardial infarction. *J Accid Emerg Med.* 1997;14:2-9.

35. Orr RK, Porter D, Hartman D. Ultrasonography to evaluate adults for appendicitis: decision making based on meta-analysis and probabilistic reasoning. *Acad Emerg Med.* 1995;2:644-650.

36. Bulloch B, Craig WR, Klassen TP. The use of antibiotics to prevent serious sequelae in children at risk for occult bacteremia: a meta-analysis. *Acad Emerg Med.* 1997;4:679-683.

37. Leydon GM, Lawrenson R, Meakin R, et al. The cost of alternative models of care for primary care patients attending accident and emergency departments: a systematic review. *J Accid Emerg Med.* 1998;15:77-83.

38. Markert RJ, Walley ME, Guttman TG, et al. A pooled analysis of the Ottawa Ankle rules used on adults in the ED. *Am J Emerg Med.* 1998;16:564-567.

39. Rothrock SG, Green SM, Harper MB, et al. Parenteral vs oral antibiotics in the prevention of serious bacterial infections in children with *Streptococcus pneumoniae* occult bacteremia: a meta-analysis. *Acad Emerg Med.* 1998;5:599-606.

40. Pearl WS, Todd KH. Ultrasonography for the initial evaluation of blunt abdominal trauma: a review of prospective trials. *Ann Emerg Med.* 1996;27:353-361.

41. Wilson A, Henry DA. Principles behind practice: meta-analysis. Part 2: assessing the quality of published meta-analyses. *Med J Aust.* 1992;156:173-187.

42. 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;352:609-613.

43. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol.* 1996;49:235-243.

44. Lecky FE, Little RA, Brennan P. The use and misuse of meta-analysis. *J Accid Emerg Med.* 1996;13:373-378.

45. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet.* 1999;354:1896-1900.

46. Rowe BH, Alderson P. The Cochrane Library: a resource for clinical problem solving in emergency medicine. *Ann Emerg Med.* 1999;34:86-90.

47. Gallagher EJ. Systematic reviews. *Acad Emerg Med.* 1999;6:1255-1260.

48. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J.* 1988;138:697-703.

APPENDIX.

OQAQ: Reviewing a Review Form.

Complete form for each study. Reviewer Initials: _____ Study ID: _____

	Quality features	1	2*	3
1	Were the search methods used to find evidence on the primary question(s) stated?	No	Partially	Yes
2	Was the search for evidence reasonably comprehensive?	No	Can't tell	Yes
3	Were the criteria used for deciding which studies to include in the overview reported?	No	Partially	Yes
4	Was bias in the selection of studies avoided?	No	Can't tell	Yes
5	Were the criteria used for assessing the validity of the included studies reported?	No	Partially	Yes
6	Was the validity of all the studies referred to in the text assessed using appropriate criteria?	No	Can't tell	Yes
7	Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	No	Partially	Yes
8	Were the findings of the relevant studies combined appropriately relative to the primary question of the overview?	No	Can't tell	Yes
9	Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?	No	Partially	Yes
10	How would you rate the scientific quality of this review?			

Flaws						
Extensive				Minor		
	Major				Minimal	
1	2	3	4	5	6	7

Note: If the methods that were used are reported incompletely relative to a specific item, score that item as "partially." Similarly, if there is no information provided regarding what was done relative to a particular question, score it as "can't tell," unless there is information in the overview to suggest either that the criterion was or was not met.

For Question 8, if no attempt has been made to combine findings, and no statement is made regarding the inappropriateness of combining findings, check "no." If a summary (general) estimate is given anywhere in the abstract, the discussion, or the summary section of the paper, and it is not reported how that estimate was derived, mark "no," even if there is a statement regarding the limitations of combining the findings of the studies reviewed. If in doubt, mark "can't tell."

For an overview to be scored as "yes" on Question 9, data (not just citations) must be reported that support the main conclusions regarding the primary question(s) that the overview addresses.

The score for Question 10, the overall scientific quality, should be based on your answers to the first 9 questions. The following guidelines can be used to assist with deriving a summary score: if the "can't tell" option is used one or more times on the preceding questions, a review is likely to have minor flaws at best, and it is difficult to rule out major flows (ie, a score ≤4). If the "no" option is used on Questions 2, 4, 6, or 8, the review is likely to have major flaws (ie, a score of ≤3, depending on the number and degree of the flaws).

*A score of 2 for any of the first 9 questions is given if the reviewer's response is "partially" or "can't tell."