The Quality of Medical Record Review Studies in the International Emergency Medicine Literature

Diana Badcock, MBBS
Anne-Maree Kelly, MD
Debra Kerr, BN, MBL
Tom Reade, MBBS

From the Department of Emergency Medicine (Badcock) and the Joseph Epstein Centre for Emergency Medicine Research (Kelly, Kerr), Western Hospital, Footscray, Victoria, Australia; the University of Melbourne (Kelly, Kerr), Melbourne, Victoria, Australia; and Sunshine Hospital (Reade), St. Albans, Victoria, Australia.

Study objective: We assess the methodologic quality of studies using medical record review methodology in 4 international emergency medicine journals. A secondary aim was to compare methodology quality among these journals and across years.

Methods: This was an observational study of articles whose main methodology was medical record review published in *Academic Emergency Medicine (AEM)*, *Annals of Emergency Medicine (Annals)*, *Emergency Medicine Journal (EMJ)*, and *Emergency Medicine Australasia (EMA)* between January 2002 and May 2004. Eligible articles were reviewed for reporting of a clear hypothesis or objective, training of abstractors, defined inclusion and exclusion criteria, use of a standard abstraction form, definition of important variables, monitoring of abstractor performance, blinding of abstractors to study hypothesis, reporting of interrater reliability, sample size or power calculation, reporting of ethics approval or waiver, and disclosure of funding source. The primary outcome was the proportion of articles meeting each criterion. Secondary outcomes were comparison of the proportions of articles meeting each criterion among journals and by years.

Results: One hundred seven articles were analyzed; 31 were published in *AEM*, 29 in *Annals*, 29 in *EMJ*, and 18 in *EMA*. A clear aim was reported in 93% of articles, standardized abstraction forms were reported in 51%, interrater reliability was reported in 25%, ethics approval or waiver was reported in 68%, and sample size or power calculation was reported in 10%.

Conclusion: Adherence to the quality criteria for medical record reviews was suboptimal, and there were significant differences among journals in overall methodologic quality. [Ann Emerg Med. 2005;45:444-447.]

INTRODUCTION

Documentation of information in medical records is of variable quality. However, many emergency medicine research studies rely on information extracted from routine medical records. It has been reported that up to 25% of published emergency medicine studies use this methodology.[1]

In 1996, Gilbert et al[1] reported an analysis of medical record reviews published in 3 emergency medicine journals between 1989 and 1993 against methodologic quality indicators. Results were disappointing, particularly with respect to reporting of use of standardized abstraction forms and testing of interrater reliability. Recently, Worster et al,[2] in an analysis of 20 articles each from 3 emergency medicine journals, reported minimal improvement in the quality of reporting of study methods. Both of these studies, however, were confined to emergency medicine journals published in North America.

The aim of this study was to assess the quality of methodologic reporting for studies using medical record review methodology in 4 international emergency medicine journals. A secondary aim was to compare methodology quality among these journals and across years.

METHODS
Study Design

This was an observational study of manuscripts using medical record review as their main methodology published in *Academic Emergency Medicine (AEM)* and *Annals of Emergency Medicine (Annals)*, published in North America, *Emergency Medicine Journal (EMJ)*, published in the United Kingdom, and

Editor's Capsule Summary

What is already known on this topic

When 3 emergency medicine journals were studied 10 years ago, articles that used medical record review had poor compliance with 8 proposed quality indicators for such methods.

What question this study addressed

Has the compliance with quality indicators increased in the past 10 years? Is there heterogeneity of compliance among emergency medicine journals?

What this study adds to our knowledge

The proportion of published studies that use medical record review has remained relatively constant. There has been modest improvement in compliance with certain quality indicators and little improvement for others. There is heterogeneity of compliance among journals.

How this might change clinical practice

This will not change clinical practice. Authors and journal editors may wish to examine the proposed criteria to determine whether their research and research publications should be structured to comply with them.

Emergency Medicine Australasia (formerly *Emergency Medicine*) *(EMA)*, published in Australia. These were chosen to give broad international representation.

Data Collection and Processing

Two researchers (DK, AMK) independently searched issues of these journals between January 2002 and May 2004 for articles that used medical record review methodology as the main source of data collection. We included studies abstracting data from clinical records, including out-of-hospital clinical records. We specifically excluded small case series and analyses of prospectively collected data or administrative databases. Where both researchers agreed about eligibility, the article was included. Where there was disagreement, a decision about inclusion was made by consensus. Eligibility was determined before evaluation of methodology.

Eligible articles were reviewed independently by 2 researchers (DB, TR). These researchers were blinded to the journal of publication by removing headers and other identifiable text. They were not blinded to the study objective. Data were collected using an explicit data collection form. No specific training was given. Each article was rated for documentation of the 8 criteria used by Gilbert et al[1] that included presence of a clear hypothesis or objective, training of abstractors, defined inclusion and exclusion criteria, use of a standard abstraction form, definition of important variables, monitoring of abstractor performance, blinding of abstractors to study hypothesis, and testing of interrater reliability. In addition, articles were

rated for documentation of sample size or power calculation, the reporting of ethics approval or waiver, and disclosure of funding source. Rating categories used in the assessment were "adhered to," "partly addressed," and "absent." When the reviewers disagreed on the rating category, the article was evaluated by a third abstractor (DK). A rating assigned by 2 of the 3 independent raters was accepted as the "true" rating. If disagreement persisted, a rating on the item was reached by consensus. Interrater reliability for the 2 principal abstractors is reported as the observed proportion of agreement in classification for the items "clear hypothesis/objective," "defined inclusion/exclusion criteria," "use of standard abstraction form," "interrater reliability reported," and "ethics approval/waiver reported."

Outcome Measures

Our primary outcome was the overall proportion of articles meeting each criterion. Secondary outcomes were comparison of the proportions of articles meeting each criterion among journals and by year.

Primary Data Analysis

Descriptive analysis was used. The study was not funded. Ethics committee approval was not required in the absence of patient-level data.

RESULTS

One hundred fourteen articles were considered for inclusion. Researchers agreed without discussion that 100 were eligible, 7 were included after consensus discussion, and 7 were excluded after consensus discussion, resulting in a sample of 107. Thirty-one were published in *AEM*, 29 in *Annals*, 29 in *EMJ*, and 18 in *EMA*. Thirty-four were published in 2002, 51 in 2003, and 22 in 2004.

Interrater agreement was 65% for definition of inclusion and exclusion criteria, 79% for use of a standard abstraction form, 85% for a clear statement of hypothesis or objective, 86% for reporting of ethics approval or waiver, and 91% for reporting of interrater reliability.

There was wide variability (0% to 100%) in each journal's compliance with each of the 11 criteria (Table 1). Analysis by year is shown in Table 2.

LIMITATIONS

There are some limitations that should be considered when our results are interpreted. We looked at only 4 English-language journals. Our results may not be generalizable to other journals that may have different submission requirements or publish articles from different research cultures. Authors may have included information about particular criteria in previous drafts, cover letters, or submission forms, but it may not have been published. The abstractors in this study were not blinded to the study objective, and that may have introduced bias. The sample size is relatively small, with 18 articles in the smallest journal group. Although we attempted to blind abstractors to

Table 1. Proportion of articles for which criteria were adhered to, overall and by journal (January 2002 to May 2004).

Criterion	Overall	AEM	Annals	EMJ	EMA
Articles, No.	108	31	29	29	18
Clear hypothesis or aim, %*	93	97	90	90	100
Training of abstractors, %*	22	29	45	3	6
Defined inclusion and exclusion criteria, %*	85	100	97	62	78
Use of standard abstraction form, %*	51	45	70	34	61
Definition of important variables, %*	68	90	86	24	72
Monitoring of abstractor performance, %*	30	29	62	3	22
Blinding of abstractors, %*	7	3	21	3	0
Interrater reliability reported, %*	28	29	52	7	22
Sample size or power calculation, %	10	16	7	7	11
Ethics approval or waiver, %	68	100	93	31	33
Funding source disclosed, %	45	29	69	65	0

*Criteria identified by Gilbert et al.[1]

Table 2. Proportion of articles for which criteria were adhered to, overall and by year.

Criterion	Overall	2002 (N=34)	2003 (N=51)	2004 (N=22)
Clear hypothesis or aim, %	93	91	94	95
Training of abstractors, %	22	15	25	27
Defined inclusion and exclusion criteria, %	85	91	78	91
Use of standard abstraction form, %	51	50	45	68
Definition of important variables, %	68	84	63	68
Monitoring of abstractor performance, %	30	35	29	23
Blinding of abstractors, %	7	3	8	14
Interrater reliability reported, %	28	24	27	36
Sample size or power calculation, %	10	9	10	14
Ethics approval or waiver, %	68	74	63	73
Funding source disclosed, %	45	29	49	59

journal, this may have been limited or unsuccessful because of the different formats of the journals studied. This study was not designed to address whether medical record review methodology was appropriate for the research questions being investigated. Finally, the criteria used to score the quality of these articles have never been validated.

DISCUSSION

Medical records are intended to document a clinical patient encounter. In truth, they are interpretations of clinical scenarios recorded by different observers who choose to record what they think is relevant or important. Missing data are common. Medical records are commonly in free-text format and often written by hand, adding the problems of legibility and interpretation.[3] Despite these weaknesses, medical record review studies may be an appropriate method as pilot studies to inform planning for prospective trials, as quality assurance activities, in determining patterns of disease throughout prolonged periods, and sometimes to investigate questions that are difficult to answer in prospective trials (eg, the effects of exposures to which patients cannot be randomized, where resources and time preclude prospective studies).[3] However, their validity rests in the quality of their methodology. A poorly conducted medical record review becomes little more than a selected case series, and the biases introduced undermine the utility of any results.

There have been a number of articles describing how to conduct a quality medical record review.[1,3] Despite these, questions persist about the quality of published research using this methodology.[2] Our study suggests that medical record review methodology as reported in international emergency medicine journals remains suboptimal. Although there have been improvements in the proportion of studies reporting use of standardized data abstraction forms, monitoring of abstractor

performance, and testing of interrater reliability since the report by Gilbert et al,[1] the actual proportions for most criteria are disappointing (Table 3). These findings are similar to those of Worster and Haines.[3]

Reasons for suboptimal performance may be research related and journal related. Research-related issues include research training and resource issues. In particular, it would appear that many medical record reviews are devised and performed by a single researcher or small group without any financial support. They perform the roles of study designer, case identifier, data abstractor, data analyst, and author. Blinding is difficult, and they monitor their own performance. Additionally, for sole researchers, recruiting a colleague to undertake data collection for interrater reliability may be logistically difficult. Journal-related issues include the standards set by journals in selection and their publication practices. For example, journals may require information about funding or conflicts of interest in a cover letter or submission form but may not publish this information.

It is interesting that the 2 North American journals had a much higher proportion of studies reporting ethics approval or waiver than the other journals, which may reflect differences in the types of studies requiring ethics approval in different countries or journal publication practices.

The differences in methodologic quality among journals are interesting. Possible explanations are different standards set in the manuscript submission and selection process, higher submission rates making selection more competitive, different editorial processes, variation in values and emphasis among journals, and international differences in research cultures. For example, at the time of the study *EMA* did not require disclosure of funding source for submissions. In Australasia, a contributing factor might also be that residents cannot graduate from training programs until completion of a research

Table 3. Comparison with findings of similar studies.

Criterion	This Study	Gilbert et al[1]	Worster et al[2]
Clear hypothesis or aim, %	93	NR	NR
Training of abstractors, %	22	17.6	NR
Defined inclusion and exclusion criteria, %	85	98.4	NR
Use of standard abstraction form, %	51	10.7	58.3
Definition of important variables, %	68	73.4	NR
Monitoring of abstractor performance, %	30	4.1	18.3
Blinding of abstractors, %	7	3.3	NR
Interrater reliability reported, %	28	0.4	11.7

NR, Not reported.

component that requires publication or presentation of a study. The favored methodology for this is medical record review, so *EMA* may have a higher proportion of studies using this methodology performed by relatively inexperienced researchers. Our study was unable to quantify this.

The deficiencies being identified, strategies for improvement are needed. These strategies should include educational and mentoring initiatives for residents and faculty and dissemination of quality standards. Journals can also play a role by making explicit minimum quality requirements for documentation of medical record review studies, similar to what has been done for randomized controlled trials,[4] and by providing constructive feedback to authors of studies on methodology improvement.

In retrospect, it would have been interesting to include a larger sample of articles to give more validity to among-journal comparisons and a broader range of journals, perhaps representing Europe and Asia.

In summary, adherence to the quality criteria for medical record review was suboptimal, although there has been some improvement in some criteria since previous studies. There were important differences among journals in methodologic quality. Strategies to improve quality in medical record review studies may include education and mentoring of researchers and standard-setting and constructive feedback by journals.

Address for correspondence: Anne-Maree Kelly, MD, Department of Emergency Medicine, Western Hospital, Footscray 3011, Victoria, Australia; 03-8345-6315, fax 03-9318-4790; E-mail Anne-Maree.Kelly@wh.org.au.

REFERENCES
1. Gilbert EH, Lowenstein SR, Koziol-McLian J, et al. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med.* 1996;27:305-308.
2. Worster A, Bledsoe RD, Cleve P, et al. Reassessing the methods of medical record review studies in emergency medicine research ten years later [abstract]. *Acad Emerg Med.* 2004;11:467.
3. Worster A, Haines T. Advanced statistics: understanding medical record review (MRR) studies. *Acad Emerg Med.* 2004;11:187-192.
4. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* 2001;357:1191-1194.