

Chart Reviews In Emergency Medicine Research: Where Are The Methods?

From the Colorado Emergency
Medicine Research Center, Division
of Emergency Medicine*, and the
Division of Internal Medicine†,
University of Colorado Health
Sciences Center, Denver, Colorado.

Received for publication
March 8, 1995. Revision received
August 7, 1995. Accepted for
publication August 23, 1995.

Copyright © by the American College
of Emergency Physicians.

Eric H Gilbert, MD*
Steven R Lowenstein, MD, MPH*
Jane Koziol-McLain, RN, MS*
Diane C Barta, MD*
John Steiner, MD, MPH†

Study objective: Medical chart reviews are often used in emergency medicine research. However, the reliability of data abstracted by chart reviews is seldom examined critically. The objective of this investigation was to determine the proportion of emergency medicine research articles that use data from chart reviews and the proportions that report methods of case selection, abstractor training, monitoring and blinding, and interrater agreement.

Methods: Research articles published in three emergency medicine journals from January 1989 through December 1993 were identified. The articles that used chart reviews were analyzed.

Results: Of 986 original research articles that were identified, 244 (25%; 95% confidence interval [CI], 22% to 28%) relied on chart reviews. Inclusion criteria were described in 98% (95% CI, 96% to 99%), and 73% (95% CI, 67% to 79%) defined the variables being analyzed. Other methods were seldom mentioned: abstractor training, 18% (95% CI, 13% to 23%); standardized abstraction forms, 11% (95% CI, 7% to 15%); periodic abstractor monitoring, 4% (95% CI, 2% to 7%); and abstractor blinding to study hypotheses, 3% (95% CI, 1% to 6%). Interrater reliability was mentioned in 5% (95% CI, 3% to 9%) and tested statistically in .4% (95% CI, 0% to 2%). A 15% random sample of articles was reassessed by a second investigator; interrater agreement was high for all eight criteria.

Conclusion: Chart review is a common method of data collection in emergency medicine research. Yet, information about the quality of the data is usually lacking. Chart reviews should be held to higher methodologic standards, or the conclusions of these studies may be in error.

[Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J: Chart reviews in emergency medicine research: Where are the methods? *Ann Emerg Med* March 1996;27:305-308.]

INTRODUCTION

Medical record reviews are used frequently to generate original research in emergency medicine. However, although medical records contain important clinical information, they are not produced for research purposes.¹ There are no universally-accepted criteria for a “well-conducted” chart review, but eight methodologic strategies (Figure 1) may enhance the validity, reproducibility, and overall quality of data collected from clinical records.^{1,2}

The goals of this study were to determine the proportion of original research articles published in emergency medicine journals that rely on retrospective chart reviews to obtain most or all of their data and to determine the proportion of chart review articles that adhere to the eight strategies.

MATERIALS AND METHODS

Original research articles published between January 1, 1989, and December 31, 1993, were retrieved from the *American Journal of Emergency Medicine*, *Annals of Emergency Medicine*, and *Journal of Emergency Medicine*. Case reports,

letters, editorials, subject reviews, metaanalyses, special theme papers, and symposium proceedings were excluded. Those of the original research articles that relied solely or mostly on data from medical records to answer the questions posed by the study were judged to be “chart review” papers. “Medical records” included emergency department or other outpatient records, paramedic reports, inpatient charts, nursing notes, and other clinical records. Studies relying on death certificates, coroners’ reports, or other public records, and all studies based on animal or laboratory investigations, were excluded.

Each chart review article was analyzed in detail for the presence or absence of methodologic standards for chart review listed in the Table. For each criterion, a rating of “Yes” or “No” was assigned. Credit was given if the authors mentioned the methodologic standard, whether or not details were provided. The articles were selected and reviewed critically by two of the authors (EHG, DCB), who were trained for this investigation. The entire research team met frequently during the study to resolve disputes and maintain consistency in selection and critical review of all articles. Standardized abstraction forms were used.

The proportion of all original research articles that relied on chart reviews and the proportions of chart review articles adhering to each of the eight criteria were determined. The results are presented as percentages with 95% confidence intervals (CIs).

A second investigator performed a blinded critical review of a random sample of 15% of the articles to determine reliability. The κ -statistic was used to measure interrater agreement.

A preliminary study reported that 4% to 35% of published medical record reviews in three internal medicine

Figure 1.
Strategies to improve accuracy and minimize inconsistencies in medical chart reviews.

Training	Train chart abstractors to perform their jobs. Describe the qualifications and training of the chart abstractors. Ideally, train abstractors before the study starts, using a set of “practice” medical records.
Case selection	Use explicit protocols and describe the criteria for case selection or exclusion.
Definition of variables	Define important variables precisely.
Abstraction forms	Use standardized abstraction forms to guide data collection. Ensure uniform handling of data that is conflicting, ambiguous, missing, or unknown.
Meetings	Hold periodic meetings with chart abstractors and study coordinators to resolve disputes and review coding rules.
Monitoring	Monitor the performance of the chart abstractors.
Blinding	Blind chart reviewers to the etiologic relation being studied or the hypotheses being tested. If groups of patients are to be compared, the abstractor should be blinded to the patient’s group assignment.
Testing of interrater agreement	A second reviewer should reabstract a sample of charts, blinded to the information obtained by the first correlation reviewer. Report a κ -statistic, intraclass coefficient, or other measure of agreement to assess interrater reliability of the data.

Table.
Adherence to methodologic standards in 244 published emergency medicine chart reviews.

Adherence	Percent	95% Confidence Interval
Abstractors trained	17.6	13-23
Inclusion/exclusion criteria described	98.4	96-99
Important variables defined	73.4	67-79
Standardized abstraction forms used	10.7	7-15
Abstractors’ performance monitored	4.1	2-7
Abstractors blinded to study objective and patient assignment	3.3	1-6
Interrater reliability mentioned	5.0	1-6
Interrater agreement tested	.4	0-2

journals adhered to various methodologic standards for chart reviews.³ A 35% adherence rate was assumed for this study, and it was determined that 200 chart review articles would be needed to provide adequate precision (95% CI limits of $\pm 6.6\%$).

RESULTS

During the 5-year period ending December 31, 1993, approximately 3,000 articles of all types were published in the three peer-reviewed emergency medicine journals. Of these, 986 were original research articles, among which chart reviews supplied most or all of the important data in 244 (25%; 95% CI, 22% to 28%). The number of chart review articles each year varied from 38 (16%) in 1991 to 61 (25%) in 1990. The majority (65.6%) of chart review articles appeared in *Annals of Emergency Medicine*.

The Table illustrates the proportion of articles that adhered to each chart abstraction standard. Most articles (98%; 95% CI, 96% to 99%) mentioned inclusion and exclusion criteria, and most defined at least one of the variables that were being analyzed (73%; 95% CI, 67% to 79%). However, compliance with the other chart review procedures was poor. Interrater reliability was mentioned in only 12 articles (5%; 95% CI, 1% to 6%), and a numeric measure of interrater agreement was reported in only 1 (.4%; 95% CI, 0% to 2%).

The data were analyzed to determine whether adherence varied across the three journals or over time. There

were no statistically significant differences among the journals, and there were no significant temporal trends.

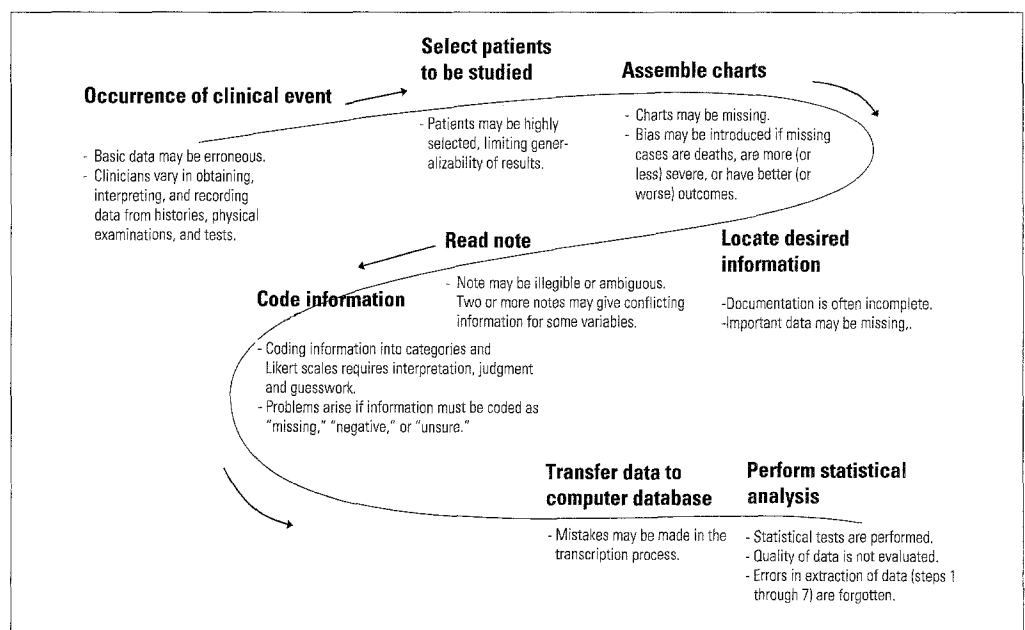
A second investigator analyzed 37 articles selected at random. The percent agreement between the two abstractors was 89% or higher for seven of the methodologic standards; the agreement was 68% for variable definitions. However, the κ -value varied widely, because the prevalence of negative answers (nonadherence to each standard) was much greater than the prevalence of positive answers. This highly unbalanced distribution of responses left little room for agreement beyond chance, imposing a high penalty on the κ -statistic.⁴

DISCUSSION

Between 1989 and 1993, one fourth of all scientific investigations published in emergency medicine journals were medical chart reviews. However, as the present investigation indicates, most chart reviews lack sound methods. Frequently, information is collected by poorly trained, unmonitored personnel, who may be aware of the study hypotheses and group or treatment assignments. Most chart reviews do not report the reproducibility of their data. Chart reviews in emergency medicine research may not always be credible sources of information.

For many reasons, medical records may not be suitable as sources of scientific data.⁵ In clinical practice, there is often poor agreement among health professionals in interpreting histories, physical signs of disease, diagnostic

Figure 2.
Process of transforming clinical record information into "hard" data is subject to error.



tests, and other clinical information.⁶⁻⁹ Clinicians also vary in the detail and accuracy with which they record their observations.⁵ In a recent study of 109 trauma patients, Schwartz et al¹⁰ demonstrated generally poor agreement between information contained in medical records and that gained from a personal interview with the patient. In one study in an ambulatory care setting, 27% of charts contained no chief complaint.¹¹ Other studies have demonstrated that 20% to 50% of abnormal laboratory results are never entered into patients' medical records.¹²

Errors, inconsistencies, and omissions in medical records are compounded when information is extracted during a scientific investigation (Figure 2). Common sources of error in chart abstraction include (1) missing charts; (2) inability to locate needed information; (3) multiple conflicting entries; (4) chart entries that are vague, incomplete, or illegible; (5) inconsistent coding of data into categories; (6) handling of uncertain or missing data; and (7) mistakes in transcription of information from charts to a computer database. In addition, biased abstraction can occur if chart abstractors are too well informed of the study hypotheses or expectations.^{2,5}

Because of errors and idiosyncrasies in the reading, interpreting, coding, and transcribing of data, information taken from medical records is often unreliable: if two abstractors were to look for the same data in the same medical record, the results often would disagree. Recent studies have demonstrated poor interrater reliability for key emergency medicine and trauma variables such as the Abbreviated Injury Scale, electrocardiography results, medication histories, preventable deaths, procedures performed, Glasgow Coma Scale scores, standard coded discharge diagnoses, "outcome" and "appropriateness of care" measures, and adverse medical events.^{7,13-16}

Medical records are informal diaries of observations, impressions, and hunches. They contain mostly verbal descriptions of people and events, and translation of these verbal descriptions into hard, quantitative data is fraught with error.⁵ Currently, in emergency medicine research there is poor adherence to recognized chart review standards. We hope that the results of this study will increase awareness by investigators and journal readers of the problems inherent in data obtained from medical charts and encourage investigators to report measures of data reliability. Medical record reviews, like laboratory studies, should be held to high methodologic standards. Fifteen years ago, Feinstein pointed out, "The tactics of extracting data from medical records . . . [are] performed according to the laws of *laissez faire*: the investigator usually chooses the records

and removes the data in whatever manner he wishes, and he seldom reports specific details of his methods."¹⁷

REFERENCES

- Boyd NF, Pater JL, Ginsburg AD, et al: Observer variation in the classification of the information from medical records. *J Chron Dis* 1979;32:327-332.
- Horwitz RJ, Yu EC: Assessing the reliability of epidemiologic data obtained from medical records. *J Chron Dis* 1984;37:825-831.
- Leininger L, Harris R, Carey T: An analysis of the quality of medical record reviews in general medicine journals. *Clin Res* 1992;40:560.
- Feinstein AR, Cicchetti DV: High agreement but low Kappa: I. The problem of two paradoxes. *J Clin Epidemiol* 1990;43:543-549.
- Feinstein AR, Pritchett JA, Schimff CR: The epidemiology of cancer therapy: III. The management of imperfect data. *Arch Intern Med* 1969;123:448-461.
- Sackett DL, Haynes RB, Guyatt GH, et al (eds): *Clinical Epidemiology: A Basic Science for Epidemiology*, ed 2. Boston: Little, Brown, 1991.
- Koran LM: The reliability of clinical methods, data and judgements I and II. *N Engl J Med* 1975;293:642-646,695-701.
- Department of Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario: Clinical disagreement: I. How often it occurs and why. II. How to avoid it and how to learn from one's mistakes. *Can Med Assoc J* 1980;123:499-504,613-617.
- Elmore JG, Feinstein AR: A bibliography of publications on observer variability (final installment). *J Clin Epidemiol* 1992;45:567-580.
- Schwartz RJ, Boiscneau D, Jacobs LM: The quantity of cause-of-injury information documented on the medical record: An appeal for injury prevention. *Acad Emerg Med* 1995;2:98-103.
- Freidin RB, Goldman L, Cecil RR: Patient-physician concordance in problem identification in the primary care setting. *Ann Intern Med* 1980;93:490-493.
- Burnum JF: The misinformation era: The fall of the medical record. *Ann Intern Med* 1989;110:482-484.
- MacKenzie EF, Steinwachs DM, Bone LR, et al: Inter-rater reliability of preventable death judgments. *J Trauma* 1992;33:292-302.
- Rowley G, Fielding K: Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 1991;337:535-538.
- Menegazzi JJ, Davis EA, Sucov AN, et al: Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. *J Trauma* 1993;34:46-48.
- Smith SM, Colwell LS, Sniezek JE: An evaluation of external cause-of-injury codes using hospital records from the Indian Health Service, 1985. *Am J Public Health* 1990;80:279-281.
- Feinstein AR, Pritchett JA, Schimff CR: The epidemiology of cancer therapy: IV. The extraction of data from medical records. *Arch Intern Med* 1969;123:571-590.

Reprint no. 47/1/69724

Address for reprints:

Steven R Lowenstein, MD, MPH
 Division of Emergency Medicine
 University of Colorado Health Sciences Center
 4200 East Ninth Avenue, Box B211
 Denver, Colorado 80262
 303-270-8711
 Fax 303-270-5614