Medical Record Review Studies
an Overview

Andrew Worster, MD, MSc

Research Director, Department of Emergency Medicine, Hamilton Health Sciences & McMaster University, Hamilton, ON, Canada

Ted Haines, MD, MSc

Faculty of Health Sciences, McMaster University Health Sciences Centre, Room 3H54 Hamilton, ON, Canada

Correspondence:
Andrew Worster, MD, MSc
Research Director
Department of Emergency Medicine
Hamilton Health Sciences & McMaster University
237 Barton Street East
Hamilton, ON, Canada
L8N 3Z5
Tel: (905) 527-4322 x 46997
Fax: (905) 527-7051
Email: worster@hhsc.ca

MeSH Terms: Medical Records
Retrospective Studies
Research Design
Evidence Based Medicine

Competing Interest: none declared

Funding: none

Abstract

The medical record review is one of the most common types of observational studies reported in the emergency medicine literature. Despite the popularity of these studies however, there are no universally accepted criteria for evaluating, reporting or conducting them. The objectives of this article are 1. to discuss the elements of medical record review studies and 2. to provide the reader with critical appraisal approaches.

Introduction

"Two years ago, non-contrast, helical computed tomography scanning (NHCT) of the abdomen was introduced in your hospital for the investigation of suspected acute urolithiasis. However, clinicians have continued to use intravenous pyelography (IVP) for this presentation, as well. You are interested in whether this change could have resulted in reduction in length of stay in the emergency department for patients investigated with CT compared to IVP. You find a paper that addressed this issue, using an analysis of medical records, but you wonder whether this kind of study can provide results that can be relied on."

Of the different types of observational studies, the medical record review (MRR) is one of the most popular amongst emergency medicine researchers. In fact, during a recent five-year period it was demonstrated that 25% of all scientific studies published in emergency medical journals were MRR studies (1).

Randomized controlled trials and systematic reviews are subject to standardized reporting in many medical journals through the use of the CONSORT and QUORUM statements respectively. Approaches to critical appraisal of studies of a number of different designs have been published in a 'how to use' series of articles (2-4). However, despite the popularity of MRR's, there are no 'universally-accepted criteria' for evaluating, reporting or conducting them (1).

The objective of this article is to discuss the elements of MRR studies and provide the reader with critical appraisal approaches for them.

Advantages and limitations

For the purpose of this paper, the term "medical record review" (MRR) refers to studies involving any form of recorded patient-focused data including physician and nursing notes, ambulance call reports and computerized medical databases. The MRR has a wide range of applications including quality assessment investigations and clinical epidemiology studies, such as studies of clinical course and prognosis.

One of the main reasons for the popularity of this method of data gathering is the accessibility of the data. Although they are labor intensive, in MRR's, a large amount of important clinical information can be found in one place, at one time and at little or no material cost to the researcher (5). One of the greatest advantages of MRR studies is that they allow the researcher to access data on events that occurred over a prolonged period of time. Therefore, they allow the researcher to answer questions of an historical nature and those that a prospective study might take too long to investigate. Although these advantages of MRR studies are self-evident, their limitations are often less obvious. These drawbacks are increased susceptibility to "bias" and to "imprecision" or "unreliability", compared to designs such as prospective studies. The results of sampling or measurement in a study are said to be biased when methodological weaknesses lead to systematically altered findings; results are imprecise (unreliable) when such weaknesses cause inconsistent findings.

As with most retrospective studies, the data of MRR's was not originally recorded for research purposes (6). Therefore, unlike prospectively collected data in which the variables are, ideally, predetermined and measurements clearly defined a priori, the MRR utilizes interpretations of different scenarios, often by different observers (1,7,8). Also, in most cases, no data quality assurance measures were in place at the time of the recording to ensure that the data were complete and accurate. These shortcomings in turn lead to a greater amount of missing data than is generally the case in prospectively planned studies and to reduced reliability and validity of the values of the recorded variables. These combined flaws can undermine the soundness of study results and conclusions.

Despite its limitations, there are research purposes and questions for which the MRR is ideal. For example, if you wanted to answer the question: "in what proportion of patients presenting to the emergency department (ED) with a chief complaint of eye problems is the visual acuity recorded on the chart?" the chart would clearly show this. In fact, in this example, the MRR is the best research method for answering the research question. It generally has strong advantages for answering research questions of a historical nature (5).

Choice of research method

The MRR is often selected as a research method because of convenience. Ideally, however, the MRR should be selected as the method for answering the research question because it is the best method, as in this visual acuity example, when validity considerations are balanced against those of cost, feasibility and state of knowledge.

The research question should reflect this. Very often there will be situations where the MRR cannot give the most valid answer compared to other designs, such as prospective ones. But, given available resources and our current level of understanding of the issue, it is the best method. For example, an MRR can give an accurate answer regarding how often VA is recorded in the ED chart but only an initial estimate of how often VA is measured in the ED. The validity of such an estimate would depend on the completeness of recording of VA measurements.

Computerized vs. paper-based medical records

Computerized databases are commonplace in emergency departments and are useful tools in retrospective clinical studies. As compared to paper-based records they: 1. are less expensive to search since no additional manpower is required to retrieve the data from a written form; 2. are more time efficient since data for a large population can be processed in a relatively short period of time; 3. provide more precise estimates when larger numbers can be analyzed. However, computerized databases are generally less accurate at the level of individual data because of the possibility of clerical error associated with the process of transcribing the data from the chart to the database (9-13). For example, Dresser demonstrated that event rates are underestimated by automated record searches unless a second record of the event exists (9). However, the risk of such errors can vary with the type of event and method of recording. Identical documentation should generally be found in cases where the event is simultaneously recorded into the chart and database. In fact, we discovered a situation in which the event was recorded in the automated record but not recorded in the paper record in 27 of 119 cases (14). It is believed that this was the result of a time delay in placing documents in patients' charts.

Record selection

The selection of records from which the data were abstracted is similar to defining the study population in prospective clinical trials. The chief complaint is a frequently used case selection criterion. Use of this variable assumes that the patient, caretaker or health care worker has correctly identified the disease process for which the patient requires help and that the chief complaint was recorded. One study conducted in an ambulatory care setting showed that no chief complaint was listed in 27% of cases (15). At our institution a review of consecutive stroke cases over a six-month period revealed that the presenting chief complaints of patients confirmed to be suffering from acute stroke included: abdominal pain; hyperglycemia; high blood pressure; shortness-of-breath; and limb pain. From these examples one can see that if the presenting chief complaint was used as the selection criterion to identify cases in a study, many cases might have been missed. On the other hand, the use of the discharge diagnosis as the selection criterion creates a risk of missing patients who may have more than one discharge diagnosis. This can occur because often only one diagnosis is listed on a database and there is no universal rule for which diagnosis should be listed first.

The criteria for case selection in studies should be clearly described to assure the reader that all, or almost all, eligible cases were selected. Similarly, the case selection criteria used in studies might result in the inadvertent inclusion of patients who are not truly cases, but if the eligibility criteria are described clearly, then the reader can evaluate these possibilities.

Sampling method

The term "sampling" refers to the method by which study cases are selected from the target population or database. A common method of sampling is to select all of the consecutive cases within a given time frame. This is a type of convenience sampling and an acceptable approach provided the period is long enough to include seasonal variations or other changes over time that are relevant to the research question (16). Less defined selection of cases is based on convenience. For example, availability of charts at a particular point in time increases the risk of unrepresentative sampling of cases.

For non-consecutive sampling, the best method of selection is "probability" or "random" sampling. Probability sampling in MRR studies provides an equal opportunity for each eligible case to be selected without bias.

Sources of data variation

It has been shown that clinical data recorded in charts are reasonably accurate in general, but the accuracy of the data is always dependent upon the recorder (5,17,18). As stated previously, most MRR data were not recorded for the purpose of research and therefore lack the strict quality criteria ideally established in prospective studies (6). The quality of the data used in any MRR study is, accordingly, only as good as the information initially recorded and the data abstraction process employed to retrieve it (5,19,20). Although MRR researchers cannot control the validity of the charted data, they can influence the accuracy with which it is abstracted. Therefore, readers should be aware of the different sources of variation in abstracted data. These have been classified into four types (20):

1. Charted reports with conflicting information;
2. Non-recording of clinical information in charts;
3. Non-availability of clinical information at the time of charting and;
4. Transcription errors - both those leading to biased data and those leading to unreliable data.

Data abstraction

As in the selection of cases, bias should be minimized in the abstraction and interpretation of data. Bias in abstracting the data can occur if the abstractors are aware of the study hypothesis (1). This is similar to the bias that can occur in prospective clinical trials when outcomes are recorded by persons who know what treatments subjects have received. Decreases in reliability in transcription can originate in differing interpretations of similar data by different abstractors and inconsistent

application of coding criteria by individual abstractors. It is therefore the obligation of the researcher to demonstrate to the readers that the data were abstracted reliably and in an unbiased manner (5). Ideally, the minimum level of inter-abstractor agreement should be 60% beyond chance agreement. Thus, this agreement should be reported as a kappa statistic, which takes chance agreement into account, rather than as percent agreement between the abstractors.

MRR's should evaluate the quality of abstraction by checking a random sample of the abstracted data against the original data. Although there are no published recommendations for what proportion of the abstracted data should be randomly checked, more is better (5).

Missing and conflicting data

A common problem with retrospective studies is missing data. In the MRR, the extent of missing data can range from partial information in charts to complete absence of entire charts. Similar to the example, mentioned above, of unlisted chief complaints, one study showed that 20-50% of abnormal laboratory results are never entered into the medical chart (15). The important question is not whether information for the study was missing, but rather how much information was missing and how did the researchers interpret and compensate for it.

As the amount of missing data increases, so does the risk of having an unrepresentative data set and thus having biased findings. Further, increasing amounts of missing data reduce the effective sample size of the study and this lowers the precision of the results - as reflected in wider confidence intervals around the summary measure.

The management of missing data in MRR's is similar in concept to intention-to-treat analysis for dropouts in prospective therapeutic trials. The onus is on the researcher to demonstrate through approaches such as sensitivity analysis that the missing values do not significantly impact on the results so as to make them questionable. For example, values covering the range of possibilities can be imputed to missing variables to determine the magnitude of potential associated effect.

As with missing data, conflicting data are a frequent feature of retrospective studies. Where two or more different versions of the same event are recorded, it is often impossible to determine after the fact which version is most accurate. In one MRR study involving multiple histories by different recorders the researchers elected to consider the history recorded by the most responsible physician as the most accurate (20). Approaches such as this are generally acceptable as long as the criteria are established a priori and applied consistently. Conflicts are an expected part of MRR studies but the number of conflicts and methods of resolution should be clearly stated.

Summary

We have discussed methodological issues and sources of error that are important to consider when reading and applying the results of MRR's to clinical practice. These have included whether the use of MRR's was appropriate given the research question, the nature of the records studied (e.g. computerized vs. paper-based), the criteria and

methods for sampling, the potential sources of error in the data and the evaluation and management of study limitations.

We recommend that readers, in reading and using MRR studies, ask the following questions related to these considerations:

1. Were the hypothesis and research question clearly stated in the introduction?
2. Was the choice of the MRR justified as the best research method?
3. Was the choice of the record database justified?
4. Were sample size considerations discussed?
5. Was the sampling strategy sound?
6. Were the main study variables well defined?
7. Were the data abstractors aware of the study hypothesis?
8. Was a Kappa statistic for reliability of abstraction reported?
9. Were the number of and reasons for missing values (per study group) provided?
10. Were the number and outcomes of data conflicts provided?
11. Was the impact of data error, missing values, and conflicts well evaluated?

While a valid way of 'scoring' the methodological strength of MRR's has not been reported, studies that emerge favorably from the scrutiny of these questions can be relied on to give a firm basis for decision-making in clinical practice.

References

1. Gilbert EH, Lowenstein SR, Kozoil-McLain J, Barta DC, Steiner J. Chart Reviews in Emergency Medicine Research: Where Are The Methods? Ann Emerg Med. 1996;27(3):305-8
2. Begg c, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: the CONSORT statement JAMA 1996;276(8):637-639
3. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354(9193):1896-900
4. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. JAMA 1993;270(17):2093-5
5. Wu L, Ashton CM. Chart review. A need for reappraisal. Eval Health Prof. 1997;20(2):146-63
6. Boyd NF, Pater JL, Ginsburg AD, Myers RE. Observer variation in the classification of the information from medical records. J Chron Dis 1979;32:327-332
7. Howie JGR. Research in General Practice. 2nd ed. 1989. Chapman and Hall Ltd. London 242p
8. Feinstein AR, Pritchett JA, Schmiff CR. The epidemiology of cancer therapy: III. The management of imperfect data. Arch Int Med 1969;123:448-461
9. Dresser MVB, Feingold Lisa, Rosenkranz SL, Coltin KL. Clinical quality measurement comparing chart review and automated methodologies. Med Care 1997;35(6):539-552
10. Lohr KN. Use of insurance claims data in measuring quality of care. Int J Technol Assess Health Care 1990;6:263
11. Gabriel SE. The sensitivity and specificity of computerized databases for the diagnosis of rheumatoid arthritis. Arthritis Rheum 1994;37:821
12. Demlo LK, Campbell PM. Reliability of information abstracted form patient's medical records. Med Care 1978;16:995
13. Roos LL, Roos NP, CaGeorge SM et al. How good are the data? Reliability of one health care databank. Med Care 1982;20:266
14. Worster A, Haines T. Does Replacing Intravenous Pyelography with Non-contrast Helical Computed Tomography Result in Benefit for Patients with Suspected Acute Urolithiasis in a Community Setting? Can Assoc Radiol j. 2002;53(3):144-8
15. Burnum JF. The misinformation era: The fall of the medical record. Ann Intern Med 1989;110:482-484
16. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing clinical research. 2 ed. Lippincott, Williams & Wilkins. Philadelphia, PA. 2001. p336
17. Ramsdell JW. Concordance of the ambulatory medical record and patients' recollections of aspects of an ambulatory new-patient visit. J Gen Int Med 1986;1:159-162
18. Wickinoff RN, Restuccia JD, Fincke BG. Concurrent application of the appropriateness evaluation protocol to acute admissions in Department of Veteran Affairs Medical Centers. Med Care 1991;29:AS64-76
19. Beard CM, Yunginer J, Reed CE, O'Connell EJ, Silverstein MD. Interobserver variability in medical record review - An epidemiological study of asthma. J Clin Epi 1992;45:1013-20
20. Horowitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. J Chron Dis. 1984;37:825-31