Performing Chart Review Studies

Editors' note: This article is the eighth in a multipart series designed to improve the knowledge base of readers, particularly novices, in the area of clinical research. A better understanding of these principles should help in reading and understanding the application of published studies. It should also help those involved in beginning their own research projects.

By definition, retrospective studies are those in which the events of interest have already occurred before the research project is begun. In other words, the interventions or exposures, the outcomes of interest, and all other relevant observations or data measurements have already occurred before the investigator begins the project. However, there are many different forms of retrospective studies, and it is unfair to lump them all together in one category. Different study designs can be used—for example, case-control studies versus retrospective cohort studies.[1]

In addition, there is a study design called "nested" studies in which the investigator retrospectively goes to an existing database and extracts the subjects of interest that are relevant to a new research project. In other words, the study subjects are found "nested" within a larger existing database. These different study designs vary substantially in many ways, including their degree of scientific validity. Therefore, it can be misleading to simply refer to all of them as retrospective or archival studies. Regardless of the study design selected, it is important to point out that all retrospective studies should have an "a priori" research question and prospectively defined study protocol before enrolling the first subject.

The most common form of retrospective research is the chart review study, a form of research that is relatively easy to perform. It is also easy to do them poorly, so these studies are frequently criticized. However, there are ways to improve the methodology and therefore the validity of these studies. This part in this series on basic research principles will address just that: how to perform chart review studies properly.

Archival Data Research

The term *archival research* refers to the use of archives or records that already exist. The medical record review, a specific type of archival data research, is the most common method to get started in clinical research. This article describes each of the elements of the medical record review used for clinical research. In a study of the emergency medicine literature, Gilbert et al[2] reviewed three different emergency medicine journals for 4 years and found that approximately 25% of their research articles used the chart review technique. That article provides a framework for reviewing manuscripts that use this technique and for planning and performing chart review studies.

Chart review studies have several advantages. First, they are relatively inexpensive to perform. The researchers simply commit to chart review time. Second, they can be accomplished quickly and using time convenient to the research team. Charts are generally available any time. Prospectively enrolled subjects are not as easily studied. Third, they do not require special laboratories or equipment. All of these factors lead to greater use of a chart review method.

Collecting data from the records retrospectively is a research technique fraught with more potential errors than is true for prospective studies. Understanding the potential pitfalls in these studies allows the investigator to attempt to address them in the research design phase. Principal to performing chart review studies is an understanding of how the information gets into the medical record. That information goes through an imprecise process to get to into the medical chart[3] and then goes through another imprecise process to be abstracted from the medical record and into the research database.[4,5] That imprecision can, and often does, lead to measurement error. If severe, those errors can completely invalidate the study results. The path the information takes to get to the research data base is summarized in Figure 1. That process comprises as many as 10 different steps. There is a potential for error at each step. It is the job of the researcher to attempt to identify and minimize these errors, to obtain the most accurate information possible. Table 1 lists an approach to performing chart review studies to maximize the validity of the results. Each step is discussed.

Limitations of Retrospective Data

Patient History

To best use the information available in the medical record, one must identify the actual events that led to the

Figure 1. The Flow of Information in Chart Review Studies

1. Actual events of interest (motor vehicle crash, syncope, chest pain)
2. Description of the events (by patient, family, EMS, police, etc.)
3. Receipt of the information (by the health care team)
4. Recording of the medical information (by the health care practitioner)
5. Coding and filing of records (by the medical records or billing units)
6. Obtaining the medical records (by the research team)
7. Interpreting the records (by the research team)
8. Abstracting data from the records (by the research team)
9. Collating, recording the data points (into database programs, by research team)
10. Analyzing, interpreting the results (by the study investigator)

Table 1. Archival Data Research: Pros and Cons

Advantages

1. Less resource-intensive than prospective research designs (faster and less expensive)
2. Can quickly evaluate a number of possible associations, which can be evaluated using more focused prospective studies
3. Can generally be performed at times of convenience

Disadvantages

1. Numerous potential sources of bias
2. Serious questions about both internal and external study validity
3. Always problems with missing data
4. Not able to establish true cause-and-effect relationships (can identify potential associations)

final chart. In the simplest case, this means that the patient, or surrogate, told something to someone, often the physician or nurse, who recorded it in the medical record. Because of the retrospective nature of chart review studies, accurately determining how the transfer of information from the patient to the record occurred and what information loss/degradation resulted is usually impossible.

Everyone is familiar with the children's game called "telephone" and how statements can change with retelling. A similar process occurs in recording the medical encounter. One study[6] showed that the physician asks enough questions to obtain only 68% of the information available about the mechanism of injury from trauma patients. Of the information obtained, only 67% of it was recorded in the medical record, so less than half of the available information actually was recorded. Differences also existed in the amount of information obtained by level of training; medical students obtained the most information from the patient but recorded the least, and attending physicians obtained the least information from the patient but recorded the most.[7-11]

Do not assume that information obtained directly from patients is always true. For example, limited patient recollection of events can result in recall bias. The longer the time period, the less a patient will remember.[12-14] Even information that we expect everyone should recall, such as children's birth weight,[15,16] allergies[17] and medication histories,[18] are frequently inaccurate. The ability of patients to recall information varies with patient characteristics, such as age,[16] medical conditions,[19] and who did the reporting.[20,21] All of these can cause errors in the data.

Documentation Process in the Medical Record

The medical record has many purposes, but research was not one of them when originally generated. Inaccuracies of the medical record are well known. The method of recording also can affect accuracy. For example, the process of dictation and transcription has been shown to introduce more inaccuracies into the medical record, such as in recording childhood immunizations.[22] However, a dictated and transcribed medical record usually contains more information than handwritten medical records.[23] Other technologies, such as voice-recognition dictation systems and other keyless entry devices, generally have improved the accuracy and completeness of documentation.[24-28] All charting technologies should be assumed to contain errors until formally evaluated for accuracy. However, electronic charts can also provide opportunities to access large databases for studies.[29]

Although not the subject of this article, it should be mentioned that registries are often used for archival studies. Disease registries (eg, cancer or trauma registries) were generally created for surveillance and epidemiologic purposes. Their development was usually not intended primarily for research or to replace the medical record. Therefore, if used for purposes other than intended, a potential for bias exists.[30-33]

Diagnoses placed on a discharge summary sheet or a billing form might have biases based on the intended use of the data. Such lists are also used by third-party payers, such as insurance companies and the Centers for Medicare and Medicaid Services. As a result, biases affect what gets documented in this list.[34] For example, The Department of Health and Human Services has agreed that hospitals can record additional diagnoses that "affect patient care, requiring clinical evaluation, therapeutic treat-

ment, diagnostic procedures, extended length of hospital stay, or increase nursing care or monitoring."[35] Other discharge diagnosis problems include indistinct coding, variable thresholds for listing chronic conditions, and reluctance for physicians to record complications.[36] Even death certificates, which are controlled by law in all 50 states, have been shown to be inaccurate. The reasons for this vary with disease entity and location but the situation is present in many nations in the world.[37,38]

Finally, the process of coding that occurs with most medical records affects subsequent database creation. This is routinely done by medical records or billing personnel.[39-41] The coding process is not done for research and can cause problems with identifying study charts. Meeting with the coders can help the researcher best identify the desired subjects.

Abstracting the Medical Record

The process of reviewing the medical record and abstracting the information to be used for research is one of the last steps in the flow of information in chart review studies. However, before any chart is abstracted for a given research project, the investigator must clearly identify the research question and case definition (which patients you are going to include and exclude), as well as all other important variables in the study. Even retrospective studies need inclusion and exclusion criteria. Once there is a definition of the study cases and variables, they should not be changed during the study. If, after reviewing the initial charts, the definitions need to be altered, the study must re-start at the beginning again with new chart abstraction. Otherwise, study patients would enter using two different criteria, which could introduce significant bias.

Keep accurate records about the charts that are available and those that are missing. Invariably, charts will be missing. If less that 5% of all the charts, it can usually be ignored as a source of bias, especially if the study is large. If 10% are missing, the results may only be 90% accurate, and an effort should be made to determine why. This could cause significant bias. Computerized logs or census data can assist in determining whether the missing charts have a common thread or are missing for a specific purpose.

No matter how diligent, there will always be some charts and individual data items that remain missing. If careful evaluation reveals that the missing items do not represent a pattern that would introduce a significant bias, a decision must be made regarding how to handle the holes in the study database. A number of potential approaches, ranging from averaging only the available data to entirely dropping that chart or group, as appropriate, are possible. Regardless of the process used, it should be established in advance wherever possible and applied consistently throughout the study. Discuss this with an experienced researcher or a statistician.

There are several important issues relating to abstractors. First, they must be qualified. Consistency and completeness are the keys to accurately reviewing charts. When possible, the person actually doing the abstraction of the medical record should not know the purpose of the research; this is called blinding and can be difficult to accomplish. It may involve lying to the abstractors. However, without it, subjective abstraction deci-

sions are prone to bias. Bias also can occur when information must be coded as "missing," "negative," or "unsure."

If the abstraction is done by multiple personnel, consistency is an issue. Differences in technique between individuals must be measured and minimized. More potential biases and errors occur as the information gets transferred from the medical research to the research database. In addition to the coding and categorical errors that can be made, simple transcription errors can always occur when entering data.

Every study should have an operations manual. At every step in the process of information flow and medical record review, the investigator should record what has been done, how, and why. Do not rely on memory. Subsequent publications should describe the study methodology in sufficient detail to allow the reader to, generally, reproduce the study themselves.

Strategies to Improve the Validity of Retrospective Studies

Although chart review studies are particularly prone to bias, not all such studies are poorly done. In fact, many are excellent. No universally accepted criteria for a "well-conducted" medical record abstraction process exist. However, there are recommended strategies to enhance the validity, reproducibility, and overall quality of data collected from clinical records.[2,3] These strategies include case selection, variable definitions, abstraction forms, training, monitoring, blinding, testing inter-rater agreement, and meetings. However, before applying these strategies or starting the data collection process, there must also be a prospective definition of the study question and the rules addressing the handling of problematic data, even though it is a "retrospective" study. Post hoc (aka "data dredging") analyses are just as scientifically invalid for retrospective research as they are for prospective studies. Together, these recommendations constitute the "10 Commandments" for properly performing chart review studies (Table 2).

Case Selection

Specific inclusion and exclusion criteria must be identified before any chart is selected for abstraction. Setting these criteria identifies the research study population. Keeping accurate counts of the subjects included and excluded is important and should be described in the manuscript results section. The case selection criteria should identify who, what, where, when, and why a patient is included or excluded.

Definition of Variables

All medical chart review studies should prospectively identify the variables that need to be abstracted from the record before performing the study. Many of these variables (eg, age, sex, death) are objective and straightforward. However, others are subjective and prone to misinterpretation. For example, what constitutes a "good outcome"? A clear definition of all study variables is necessary for accurate and consistent abstraction from the medical record. These definitions should be agreed on by all of the investigators in advance. The definitions must be taught to the abstractors. A study dictionary, containing all the defini-

Table 2. The 10 Commandments for Performing Chart Review Research

1. Prospectively (a priori) define the research question before any data collection.
2. Prospectively define the study case selection.
   - Detailed study inclusion and exclusion criteria
3. Prospectively define study variables and develop study policies.
   - These may need to be expanded or modified as the study progresses.
4. Ensure high-quality data abstractors.
   - Qualifications and training
5. Ensure consistent data recording.
   - Such as with a detailed data form or computer program
6. "Blind" the data abstractors to the study purpose (whenever possible).
7. "QA" the data collection process through periodic monitoring.
8. "QA" data processing through duplicate entry techniques.
9. Be consistent throughout.
   - If there are significant changes in any major study variables, start over from the beginning.
10. Monitor the study progress.
    - Hold periodic meetings.

tions, should be generated that serves as a reference during and after the study period. The subsequent manuscript should also include definitions of the key variables used in the study.

Abstraction Forms

The chart review methods and the abstraction process must be standardized. It is routine to use template abstraction forms. This guides data collection and ensures uniform recording of data. The form should be designed to be easy to use and have sufficient space to record all the information. It is helpful to design the form to collect data elements in the order that they might be present in the medical record.

Training

With the notable exception of medical records personnel, few of us are trained to review the medical record and abstract data and information. Having experience with medical documentation does not guarantee an ability to accurately review and interpret the medical record. It is imperative that the chart reviewers be qualified to perform the job. They should be trained by someone who is knowledgeable about the medical record, usually the principal investigator. The training should cover all of the parts of the record that need to be reviewed, identifying medical synonyms and colloquialisms that might be used, and discouraging subjective interpretations of the information during abstraction. Training can be time consum-

ing. Training medical students, for example, to properly review a record could take over 5 hours. Novice abstractors should then be given "practice" charts to review as part of the initial training. When the abstractor becomes proficient with the practice charts, they can begin abstracting the "real" charts from the study. If the abstractor is still in doubt about a given data item, they should identify the difficulty, copy the exact statements, and review with the primary investigator. If relevant, a rule or policy should be developed for dealing with such data in the future to ensure consistency.

Monitoring

The principal investigator should monitor the performance of the chart abstractors to identify any problems. Problems with incomplete or poor interpretation of the chart, taking shortcuts, and misplacing charts are unfortunate but do occur. The abstractors should be held accountable for the quality of their work. When monitoring is performed, it should be described in the methods portion of the manuscript, so the reader can understand the diligence put into the project.

Blinding

As described earlier, the chart reviewers should be blinded (not be allowed to know) to the study question and hypothesis, or the research purpose. This is not always possible, but it is worth the effort. Nonblinded review of medical records can be very problematic. Subjective bias can be very hard to control if there is not blinding. If this is not possible, an explanation should be provided in the manuscript and included in the discussion of study limitations.

Testing Inter-rater Agreement

In studies with more than one chart abstractor involved, it is important to determine whether the abstraction is being performed in a consistent manner. One approach to improve the abstraction process is to generate an example chart displaying the relevant information that can be used as a reference by the abstractors. There are ways to test the inter-rater agreement, called reliability, by having both abstractors abstract a sample of the same charts. Neither should have any prior knowledge of the information obtained from the charts before abstraction (blinded review). The abstraction results are then compared using a statistical measure of agreement, commonly a kappa statistic or intraclass correlation coefficient. It would be optimal to study the inter-rater reliability before starting the study, during the study, and at the end of the study, while blinding the process from the abstractors themselves. Another approach to maximize reliability is to use a "dual" data entry technique. This involves having every chart abstracted by two reviewers. These results are compared. Any differences are then adjudicated by a higher authority, usually the study primary investigator. This is a highly scientific approach but can double the amount of work.

Meetings

The principal investigator needs stay informed during the chart abstraction phase of the project. It is best to remedy problems early before they result in significant bias. This

requires routine scheduling of meetings with chart abstractors and study coordinators to resolve disputes and conflicts and to review coding rules. Initially, such meetings should be often, then less frequent as needed.

Conclusion

Medical records are informal collections of observations and impressions that contain both subjective and objective information obtained during the patient care process. They are not created or designed for research but frequently are used for that secondary purpose. Chart review studies are more prone to bias and other errors than is true for prospective studies. Adhering to guidelines for proper chart review technique ensures a more valid and reliable study and improves the quality of medical record review research.

References

1. Schwartz RJ, Jacobs LM. Analysis and comparison of research abstracts at MMS, 1987-1990. Air Med J 1992;11:7-11.
2. Gilbert EH, Lowenstein SR, Koziol-Mclain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: Where are the methods? Ann Emerg Med 1996;27:305-308.
3. Horwitz RI, Yu ED. Assessing the reliability of epidemiologic data obtained from medical records. J Chron Dis 1984;37:825-831.
4. Bertelsen J. Who should abstract medical records? A study of accuracy and cost. Evaluation Health Professions 1981;4:79-92.
5. Leininger L, Harris R, Carey T. An analysis of the quality of medical record reviews in general medicine journals [abstract]. Clin Res 1992;123:560.
6. Schwartz RJ, Boisoneau D, Jacobs LM, The quantity of cause-of-injury information documented on the medical record: an appeal of injury prevention. Acad Emerg Med 1995;2:98-103.
7. Agius RM, Lee RJ, Symington IS, Riddle HF, Seaton A. An audit of occupational medicine consultation records. Occup Med 1994;44:151-157.
8. Austin K. The identification of mistakes in road accident records. Part 2: Casualty variables. Accid Anal Prev 1995;27:277-282.
9. Bairstow BM, Burke V, Beilin IJ, Deutscher C. Inadequate recording of alcohol drinking, tobacco-smoking and discharge diagnosis in medical in-patients: failure to recognize risks including drug interactions. Med Educ 1993;27:518-523.
10. Kennedy GT, Stem MP, Crawford MH. Miscoding of hospital discharges as acute myocardial infarction: implications for surveillance programs aimed at elucidating trends in coronary artery disease. Am J Cardiol 1984;53:l000-1002.
11. Martie TJ, Durant H, Sealy E. Pneumonia: the quality of medical records data. Med Care 1987;25:20-24.
12. Hale WA, Delaney MJ, Cable T. Accuracy of patient recall and chart documentation of falls. J Am Board Fam Pract 1993;6:239-242.
13. Casey R, Rieckhoff M, Beebe SA, Pinto-Martin J. Obstetric and perinatal events: the accuracy of maternal report. Clin Pediatr 1992;31:200-204.
14. Tilley BC, Barnes AB, Bergstralh E, , Labarthe D, Noller KL, Colton T, et al. A comparison of pregnancy history recall and medical records: implications for retrospective studies. Am J Epidemiol 1985;121:269-281.
15. Pyles MK, Stolz HR, MacFarlane JW. The accuracy of mothers' reports on birth and developmental data. Child Dev 1935;6:165-176.
16. Oates RK, Forrest D. Reliability of mothers' reports of birth data. Aust Paediatr J 1984;20:185-186.
17. Preston SL, Briceland IL, LesarTS. Accuracy of penicillin allergy reporting. Am J Hosp Pharm 1994;51:79-84.
18. Beers MH, Munekata M, Storrie M. The accuracy of medication histories in the hospital medical records of elderly persons. J Am Geriatr Soc 1990;38:1183-1187.
19. Linet MS, Harlow SD, Mclaughlin JK, McCaffrey LD. A comparison of interview data and medical records for previous medical conditions and surgery. J Clin Epidemiol 1989;42:l207-12l3.
20. Tsubono Y, Fukao A, Hisamichi S, Hosokawa T, Sugawara N. Accuracy of self-report for stomach cancer screening. J Clin Epidemiol 1994;47:988-981.
21. Bondy ML, Strom SS, Colopy MW, Brown BW, Strong Le. Accuracy of family history of cancer obtained through interviews with relatives of patients with childhood sarcoma. J Clin Epidemiol 1994;47:89-96.
22. Wilton R, Pennisi AJ. Evaluating the accuracy of transcribed computer-stored immunization data. Pediatrics 1994;94:902-906.
23. Stueven HA, Tonsfeldt DJ, Hargarten KM, Olson DW. A dictated and transcribed medical record can be cost effective. J Am Record Assoc 1991;62:37-40.
24. Patrikas EO. Research review: use of keyless data entry in medical record departments. Top Health Inform Manage 1993;14:69-76.
25. Linn NA, Rubenstein RM, Bowler AE, Dixon JL. Improving the quality of emergency department documentation using the voice-activated word processor: interim results. Proc Annu Symp Comput Appl Med Care 1992:772-776.
26. Holbrook J, Aghababian R. A computerized audit of 15,009 emergency department records. Ann Emerg Med 1990;19:139-144.
27. Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer systems. Br Med J 1993;307:32-34.
28. Chua RV, Cordell WH, Emsting KL, Bock HC, Nyhuis AW. Accuracy of bar codes vs handwriting for recording trauma resuscitation events. Ann Emerg Med 1993;22:1545-1550.
29. Payne TH, Goroll AH, Morgan M, Barnett GO. Conducting a matched-pairs historical cohort study with a computer-based ambulatory medical record system. Comput Biomed Res 1990;23:455-472.
30. Ing RT, Baker SP, Eller JB, et al. Injury surveillance systems: strengths, weaknesses, and issues workshop. Public Health Rep 1985;100:582-586.
31. Ribbeck BM, Runge JW, Thomason MH, Baker JW. Injury surveillance: a method for recording E codes for injured emergency department patients. Ann Emerg Med 1992;21:37-40.
32. Langdorf MI, Strange G, Macneil P. Computerized tracking of emergency medicine resident clinical experience. Ann Emerg Med 1990;19:764-773.
33. Roos LL, Mustard CA, Nicol JP, McLerran DF, Malenka DJ, Young TK,et al. Registries and administrative data: organization and accuracy. Med Care 1993;31:201-212.
34. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. N Engl J Med 1988;318:352-355.
35. Coding Clinic for ICD-9-CM. Chicago, IL: American Hospital Association Division of Quality Control Management. 1990;7:13.
36. Jencks SF. Accuracy in recorded diagnoses. JAMA 1992;267:2238-2239.
37. Moussa MA, Shafie MZ, Khogali MM, el-Sayed AM, Sugathan TN, Cherian G, et al. Reliability of death certificate diagnoses. J Clin Epidemiol 1990;43:1285-1295.
38. Lapidus G, Braddock M, Schwartz R, Banco L, Jacobs LM. Accuracy of fatal motorcycle-injury reporting on death certificates. Accid Anal Prev 1994;26:535-545.
39. Bertelsen J. Who should abstract medical records? A study of accuracy and cost. Evaluation Health Professions 1981;4:79-92.
40. Dawson-Saunders B, Mast TA, Finch WT, Konrad HR, Folse JR. Content knowledge and problem-solving skill in reviewing medical charts. Med Educ 1984;18:31-35.
41. Yeoh C, Davies H. Clinical coding: completeness and accuracy when doctors take it on [see comments]. BMJ 1993;306:972.

Edward A. Panacek, MD, MPH, is professor of emergency medicine and clinical toxicology at the UC Davis Medical Center in Sacramento, California. He can be reached at eapanacek@ucdavis.edu.