ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 269–274

Journal of Clinical Epidemiology

Guidelines were developed for data collection from medical records for use in retrospective analyses

Angelique C.M. Jansen[a], Emily S. van Aalst-Cohen[a], Barbara A. Hutten[b], Harry R. Büller[a], John J.P. Kastelein[a,*], Martin H. Prins[c]

[a]*Department of Vascular Medicine, Academic Medical Center, University of Amsterdam, Room F4-159.2, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands*

[b]*Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands*

[c]*Department of Clinical Epidemiology and Medical Technology Assessment, Academic Hospital Maastricht, Maastricht, The Netherlands*

Accepted 12 July 2004

Abstract

Objective: To construct a set of guidelines for data collection from medical records.

Study Design and Setting: Retrospective analysis of clinical data is often performed by physician-scientists. In such research, the source of clinical data is the patient's medical record; however, medical records are intended for patient care and the data are not systematically recorded for research purposes. We drew on recommendations in the literature and our own experience with a retrospective cohort study that uses a DNA bank to construct guidelines for data collection from medical records.

Results: The guidelines incorporate a number of strategies for accurate data collection, which are discussed and illustrated by application.

Conclusion: With guidelines for data collection, the quality of research data is enhanced. A well-designed case record form and a handbook for standardized data collection are essential for training the data collectors and for ensuring fastidious searching of the record; however, certain kinds of information are not always well documented in patient records. Consequently, it is essential to perform a pilot study to assess the study design and to use additional questionnaires. Correct interpretation of clinical outcomes documented in the medical records often necessitates an independent adjudication committee to prevent bias in outcome definition.

Keywords: Data collection; Medical record; Retrospective; Validation studies; Methodology; Bias

1. Introduction

During recent years, biobanks of patient materials such as serum, DNA, and pathology specimens have become a rich source for scientific research. Such patient materials are stored in laboratory freezers, pending use with new diagnostic techniques when such become available—and, indeed, retrospective examination and analysis of biobank materials and other clinical data are performed increasingly by physician-scientists and epidemiologists.

In such a retrospective study, the primary source of clinical data is almost always the medical records of the participating patients; however, medical records are intended primarily for patient care and the data are not systematically recorded for research purposes. Nevertheless, retrospective studies using such data should be of high quality, without incomplete, inappropriately recorded, or missing data. In analogy, it is expected that data collection in randomized controlled trials (RCTs) is of the highest quality [1,2] as their unbiased evaluation of medical treatment has a major impact on medicine. Observational studies, such as cohort studies using patient records, likewise have a considerable impact on medical practice. In fact, such studies are performed even more often than RCTs, because it is relatively easy to collect the necessary data and the attached costs are comparatively low [3].

In the process of designing one of our current research projects, the GIRaFH study (Genetic Identification of Risk Factors in Familial Hypercholesterolemia), which uses a large DNA bank, we performed a systematic search of the published literature for the design, execution, and reporting of retrospective studies using medical records for data collection. No comprehensive guidelines were found for the execution or reporting of such studies. Therefore, we decided to

* Corresponding author. Tel.: +31-20-566-6612; fax: +31-20-566-9343.

E-mail address: e.vandongen@amc.uva.nl (J.J.P Kastelein).

0895-4356/05/$ – see front matter
doi: 10.1016/j.jclinepi.2004.07.006

develop a set of guidelines, which are presented here. These guidelines were developed drawing on recommendations from the published literature and our own experience with the GIRaFH study. Subsequently, we assessed the contribution of the constructed guidelines to the quality of the GIRaFH study and their possible implications for future research.

2. Materials and methods

2.1. Literature

The MEDLINE database for the period January 1966 through May 2004 was searched using the following key terms: medical record, chart review, retrospective study, observational study, validation studies, methodology, study design, peer review, reporting, quality management, bias, and confounding. In addition to examining several biostatistics and clinical epidemiology textbooks [4–6], we evaluated major publications (and their references) on the quality assessment of clinical research, including papers on randomized controlled trials, pharmacological studies, meta-analyses, and observational studies [1–3,7–12]. Furthermore, we evaluated recent studies using retrospectively collected data and compared their use of such data to ours [13–16].

2.2. Genetic Identification of Risk Factors in Familial Hypercholesterolemia: The GIRaFH study

Heterozygous familial hypercholesterolemia (FH) is a common (1:400) hereditary disorder of lipoprotein metabolism. Due to genetic defects in the low-density-lipoprotein receptor gene, patients suffer from severely elevated low-density lipoprotein (LDL) cholesterol levels and, as a consequence, from early atherosclerosis and premature cardiovascular disease (CVD). Although FH is a monogenic disorder, variation is observed in the severity and onset of cardiovascular symptoms. The study objective was to estimate the contribution of genetic variations to the development of CVD in a large cohort of FH patients.

A retrospective, multicenter cohort study was performed in 2,400 FH patients from lipid clinics of 27 hospitals throughout the Netherlands. These patients were randomly selected from the DNA-bank database of the Department of Vascular Medicine at the Academic Medical Center in Amsterdam, which has been appointed as the official molecular diagnostic center for nationwide FH screening in the Netherlands.

Phenotypical data were acquired by reviewing medical records by a well-trained team of 13 data collectors. Strict inclusion and exclusion criteria were applied to ensure the inclusion of definite FH patients in the study. Data were collected on demographics, classical risk factors, medication use, physical examinations, laboratory parameters, and extensive information on CVD. All patients gave informed consent and the Ethics Institutional Review Board of each participating hospital approved the protocol.

2.3. Flow of information: The data-collection process

To arrive at the present guidelines, we examined the flow of information in the data-collection process and designed strategies for accurate data collection based on the literature and our own experience. Figure 1 shows the flow of information for data gathered from patient to medical record (a), and from medical record to database (b). The figure also presents several proposed tools for consistent data collection (pilot study, case record form, handbook, questionnaire, and independent adjudication committee) and where they may play a role, as discussed below.

2.3.1. Information from patient to medical record

A medical record contains information supplied by the patient to the physician. This information is often not standardized or complete and is prone to subjectivity. For example, the patient may recall information from his or her earlier medical history incorrectly, or may report symptoms incompletely or inaccurately (e.g., gastroesophageal reflux reported as angina). Furthermore, the physician may take an incomplete history or may record information incorrectly. In addition, it must be taken into account that certain kinds of information (e.g., data on potential confounders) may be lacking in older records, without the benefit of subsequent advances in medical knowledge. For instance, homocysteine has only recently been recognized as a risk factor for CVD and may not be listed in earlier records.

When researchers refer to a medical record for research data, the patient and physician are usually not consulted. Therefore, errors occurring at the patient and physician levels are difficult to avoid. To evaluate possible errors, questionnaires may be sent to a random selection of patients and checks may be performed on the information in the medical record versus that in the questionnaire. If important differences are identified, the researchers should send questionnaires to all participating patients. To reduce possible errors, the data collector should verify any recorded information against the questionnaires in addition to original source documents such as hospital discharge reports and other physician's notes.

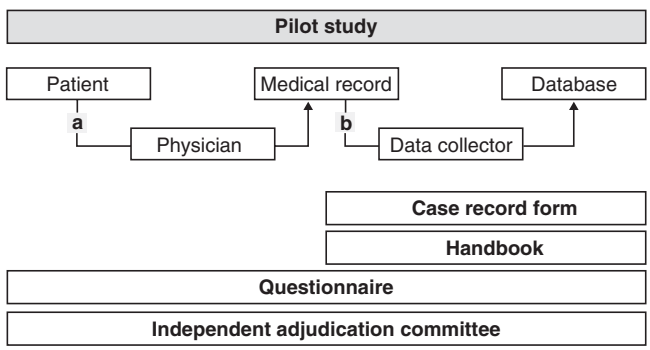

Fig. 1. Data information flows (a) from patient to medical record and (b) from medical record to database.

2.3.2. Information from medical record to database—the medical record

In general, as in a randomized controlled trial, the data collected in a retrospective cohort study should be standardized for each patient. Standardization of the data is important for the internal validity of a study, but is also essential for the reproducibility of the research data. Data collection can best be standardized by using a case record form (CRF) that is easy to complete and effective for collecting the required data. In the literature, we found useful publications on how to design an effective CRF [17,18]. Another advantage of using CRFs rather than computerized data collection is that a CRF is easier to authenticate for later referral [18].

The availability of information from the medical record is important for answering the research question. We could not find recommendations about what constitutes adequate availability. Therefore, we arbitrarily decided on good availability of data as >80% for clinical and laboratory parameters, and >95% for clinical outcomes. If the availability of data is not sufficient, the investigator should send a supplementary questionnaire to all participating patients during definitive data collection.

Incorrect transfer of data from the medical record to the CRF is a problem in large studies and when multiple data collectors are involved. To prevent this, random checks of the CRFs must be performed prior to data entry, comparing the original records to the information listed in the CRF. This can be achieved by applying the 100–20 rule: 100% of the data are checked in 20% of the CRFs, and 20% of the most essential data are checked in 100% of the CRFs [19]. Furthermore, when data entry from the CRF to the database is done by paramedical personnel, inconsistencies may be detected upon data entry and reported in queries to the principal investigator. In addition, consistency checks run on the database may reveal outlier results. All errors found must be corrected using the original medical records.

2.3.3. Information from medical record to database—the data collectors

For a large study, a team of specially trained data collectors may be effective in dealing with a large number of data in a relatively short period. Splitting such a team into smaller units for multisite data collection has been shown to be superior to using on-site data collectors [20].

To avoid bias, information about clinical outcomes should be obtained in the same way for all patients. Using CRFs and a handbook of standards minimizes information bias [4–6,12].

A handbook of standards for completing each CRF item can be helpful in attaining maximal uniformity of data records and in avoiding misinterpretation of data. For example, the handbook can provide lists of stringent criteria for inclusion and exclusion characteristics of the patients in addition to other clinical data. General instructions should include how to deal with incomplete or missing data: for instance, so-called dummy data reduce the number of incomplete data such as dates. Using such a handbook helps train the data collectors in critically examining the patient information, the physician's notes, and the physician's thought processes.

Interobserver variability poses a threat to the quality of the data collected, but a well-designed CRF and handbook can reduce this risk. In addition, interobserver discrepancies should be evaluated during data-collection training. For example, each data collector gathers data from 10 medical records onto CRFs and the principal investigator does the same. The data entered into the two sets of CRFs are subsequently entered into a data entry program (e.g., SPSS software; SPSS, Chicago, IL, USA) and compared using a computer. If necessary, further training sessions are performed. In this manner, interobserver variability can be reduced substantially.

2.4. Performing a pilot study

In our opinion, a pilot study performed prior to initiating the study is essential for reconsidering elements of the study design: the research question, the study population, the inclusion and exclusion criteria, the CRF, the accuracy of the collected data, and the availability of information from the various data sources. During the pilot study, the entire data-collection process is scrutinized, as shown in Fig. 1.

2.5. Independent adjudication committee

Consulting an independent adjudication committee is essential when there are no consensus criteria for the definition of the disease and clinical outcome of interest. In cases where the clinical outcome in question does not quite fulfill the criteria, the committee reviews all of the available data and makes a final decision regarding outcome status.

3. Results

3.1. Guidelines

We developed a set of guidelines that incorporated strategies for accurate data collection (Table 1). To illustrate the application of the proposed guidelines, we discuss three essential aspects of the GIRaFH study (inclusion and exclusion criteria, information on smoking, and information on cardiovascular events) and how these data were retrieved from the medical records.

3.2. Data collection overall

A team of 13 data collectors was involved in the present study. Therefore, interobserver variability was a real threat to the quality of the data collected. After performing the pilot study, the CRF and handbook for data collection were redesigned to reduce this variability. In addition, during the final study, interobserver discrepancies were evaluated

Table 1
Guidelines for data collection from medical records

Risks	Strategies for inconsistency reduction
a. Information flow from patient to medical record	
Patient and/or physician	
Incorrect information supplied by the patient and/or incorrectly recorded information.	Verify information against original source documents and multiple physicians' notes.
	Send additional questionnaires to a random selection of participating patients and perform checks of medical record information versus questionnaire information.
Diagnostic reports from the wrong patient	Check whether diagnostic reports correspond to the patient.
b. Information flow from medical record to database	
Medical record	
Data from medical record may not be available	Check availability in a pilot study. Send additional questionnaires to all participating patients if necessary.
Medical record may not be complete	Data collectors should cooperate closely with local physicians and other staff, who should provide other available records or sources of information.
Data collectors	
Incorrect interpretation of data in medical record	Data collectors must have a paramedical educational background.
	Consult an independent adjucation committee on definitions of disease and clinical outcome.
	Perform an interobserver study as part of data collection training.
	Supply handbook of written guidelines to data collectors defining consensus criteria for medical conditions and other rules for data collection.
Nonstandardized data collection	Create a case record form (CRF).
	Use handbook of data collection guidelines.
Unauthenticated data collection	Retain CRF as authenticated data.
	Save completed and checked CRFs as original data for later referral.
Uncertain endpoints	Form an independent adjudication committee.
Illegible physician's handwriting	Ask different data collectors to judge the same handwriting.
Inconsistencies in multicenter study with on-site data collection	Employ one team of data collectors across the different study sites.
	Perform an interobserver study between team members to control interobserver variability.
Biased data collectors	Blind the data collectors.
Uninformed data collectors	Organize regular meetings between data collectors to discuss data collected and interpretation thereof.
Incorrect transfer of data from medical record to CRF	Perform a random check of completed CRFs against original medical records.
	Perform a consistency check of final database to reveal outlier results.

For information flow, (a) and (b) refer to the matching flowchart *a* and *b* in Fig. 1.

during data-collection training. Each new data collector gathered data from 10 medical records onto CRFs and the principal investigator did the same. The data entered into the two sets of CRFs were subsequently entered into the SPSS data entry program (version 11.0) and compared using a computer. On average, 10% of the data differed between the new data collector and the principal investigator. Further training sessions were performed for all new data collectors, leading to a reduction in variability of $<1\%$.

3.3. Inclusion and exclusion criteria

DNA samples of patients with clinically suspected FH are regularly sent to our DNA laboratory from lipid clinics throughout the Netherlands for analysis of the LDL receptor gene. For the GIRaFH study, 4,000 potential patients were randomly selected from this database. Inclusion and exclusion criteria were carefully defined in the handbook as an aid to the data collectors and to avoid selection bias. Patients without FH and/or with secondary hypercholesterolemia were excluded. After these criteria were applied, a total of 2,400 patients were included in the study.

3.4. Smoking

Smoking is an important risk factor for cardiovascular disease. In the pilot study, smoking status could be retrieved from only 56% of the records. When we then adapted the final CRF to contain more specific information on lifetime smoking status, the availability increased to 68%. To achieve maximal smoking status information, questionnaires were sent to all participating patients; this further increased the information availability percentage to 88%.

The handbook contained tables useful for standardizing recorded numbers of cigarettes. For example, if the doctor recorded that the patient "only smoked at parties," the data collector standardized this into 0.5 cigarette a day. Such items were discussed by members of the study team and standardized definitions were decided by consensus.

3.5. Cardiovascular events

One of the goals of the pilot study was to assess whether the selected study population would be representative of the FH patient population at large (FH patients who are seen in outpatient clinics) and would therefore be appropriate for answering the research question. One of the important indicators was therefore the cardiovascular event rate. The pilot study population demonstrated an event rate during the observation period that was consistent with the literature on cardiovascular disease in FH patients (32% of patients had clinically manifest cardiovascular disease) [21]. The final study data exhibited similar event rates [22].

Cardiovascular events were scored using the handbook which listed recent consensus criteria for these events. Independent sources such as the physician's notes on patient symptoms, the physician's notes on possible clinical diagnoses, and the diagnostic reports were reviewed by the data collectors. This information was combined to evaluate whether or not the event fulfilled the criteria.

In case the events did not quite fulfill the criteria or if any suspect histories, symptoms, or diagnostic evaluations were discovered in the record, the case was presented to an independent adjudication committee consisting of a cardiologist, a neurologist, and a vascular surgeon, using anonymous copies of the necessary documents from the medical record. In retrospect, we proved to have interpreted nearly all events correctly prior to consulting the committee.

4. Discussion

The GIRaFH study, by nature of its design, is prone to bias [4–6,12]. By using guidelines for consistent data collection, however, we believe we have enhanced the quality of the research data.

Defining clear inclusion and exclusion criteria is essential in any research project; neglecting to do so may result in a heterogeneous patient population in which potential confounders abound. In a study using medical records as the primary data source, these criteria must be especially clear to all data collectors and must be sought for in all source documents. A clear CRF and a handbook for standardized data collection are essential for training the data collectors to apply the criteria and to stimulate fastidious searching of the record; however, certain kinds of information are not always well documented in patient records. In our experience, it is helpful to perform a pilot study. Particularly when a project is dependent solely on data from medical records, a pilot study helps to assess the availability of the necessary information, the feasibility of the project, and the quality of the CRF. Standardized data collection is further enhanced by data-collection training of data collectors to reduce interobserver variability.

Despite the fact that smoking is an important predictor of cardiovascular disease, it is often poorly documented in medical records. Underreporting of smoking could lead to incorrect estimation of the relative risk of a certain candidate gene associated with increased cardiovascular risk. It was therefore important to assess the availability of the smoking status from the medical records. During the pilot study, smoking status was available in only half of the records. Adaptation of the CRF and handbook according to recommendations made after the pilot study, as well as use of questionnaires, increased the information on smoking status; however, we must take into account that some smokers failed to return the questionnaire (nonrespondent bias), which could still lead to an incorrect estimation of the relative risk. Fortunately, the response rate was high (70%), and we therefore considered this type of bias to be minimal.

In the present study, the genetic make up of FH patients with and without cardiovascular events was compared. It was therefore important to define which patients had experienced a cardiovascular event. We hoped that by consulting an independent adjudication committee, we could avoid misinterpreting events recorded in the medical records. In the event, the number of misinterpreted events proved to be insignificant. We believe that the meticulous definition and application of cardiovascular event criteria led to this result. Consulting an independent committee did not add to the quality of the GIRaFH study but an independent adjudication committee may be important when explicit definitions of disease and clinical outcome are not available. The definition of a myocardial infarction and consequently the consistent interpretation of clinical findings and laboratory results to define the occurrence of myocardial infarction are better developed than the criteria for pancreatitis. Consulting the independent adjudication committee did, however, give an indirect indication of the quality of our data collection for our interpretation of nearly all events proved to be correct prior to consulting the committee.

The uniqueness of our guidelines becomes apparent when one reviews the available literature on retrospectively collected data from medical records. First, recent studies using such data evaluate the use of administrative or electronic databases and not medical records per se [13–16]. Second, the quality of the data is indeed questioned in these studies, but not the quality of data collection as such. The quality of the latter is questioned in the textbooks we referred to; however, no comprehensive practical solutions for quality management of this type of data collection are provided [4–6]. Third, we have created general guidelines, rather than specific guidelines for a particular type of study design. General guidelines could be referred to as a gold standard by almost any study using this type of data, thereby ensuring a certain degree of quality. Moreover, we believe that today's physician-scientist will refer more readily to a simple, easily accessible document.

It is widely recommended that research projects undergo quality review. As yet, however, there are no comprehensive guidelines for performing retrospective cohort studies using medical records—even though this type of design is increasingly used, due to the availability of large collections of

patient materials. The recommendations here are based on our own experience with the difficulties we encountered. The results do not give a definitive solution for minimizing bias during data collection from medical records, but are suggestions for quality management of data collection. We believe that attention paid to the methodology of such studies will stimulate adequate reporting. We expect that this will ultimately lead to the creation of effective checklists that are easy to use by physician-scientists.

Acknowledgments

The authors are grateful to the Netherlands Heart Foundation for their support (grant NHS 98/165). J.J.P. Kastelein is an established investigator of the Netherlands Heart Foundation (grant D039/66510). The authors wish to thank Mrs. G.E.E. van Noppen for her valuable assistance with manuscript preparation.

References

[1] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191–4.

[2] Spriet A, Dupin-Spriet T. Monitoring a clinical trial. In: Good practice of clinical drug trials. Coluzzi R, Young J, translators. 2nd ed. Basel: Karger;1997:77–91.

[3] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002;359:57–61.

[4] Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.

[5] Gail MH, Benichou J, editors. Encyclopedia of epidemiologic methods. Chichester: Wiley; 2000.

[6] Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven; 1998.

[7] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB, Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA 2000;283:2008–12.

[8] DuRant RH. Checklist for the evaluation of research articles. J Adolesc Health 1994;15:4–8.

[9] Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. JAMA 1999;281:1919–26.

[10] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896–900.

[11] Sackett DL. Bias in analytic research. J Chron Dis 1979;32:51–63.

[12] Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet 2002;359:248–52.

[13] Godley PA, Rathore SS, Kshirsagar AV, Amamoo MA, Schell MJ, Freeman J, Harris RP. Retrospective classification of prostate-specific antigen tests: differentiating screening from diagnostic clinical encounters. J Clin Epidemiol 2001;54:884–8.

[14] Breitfeld PP, Dale T, Kohne J, Hui S, Tierney WM. Accurate case finding using linked electronic clinical and administrative data at a children's hospital. J Clin Epidemiol 2001;54:1037–45.

[15] Chiò A, Ciccone G, Calvo A, Vercellino M, Di Vito N, Ghiglione P, Mutani R. Piemeonte and Valle d'Áosta register for ALS. Validity of hospital morbidity records for amyotrophic lateral sclerosis: a population-based study. J Clin Epidemiol 2002;55:723–7.

[16] Oostenbrink R, Moons KGM, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: prospects and problems. J Clin Epidemiol 2003;56:501–6.

[17] Lawrence G. Case record form design. In: Rondel RK, Varley SA, Webb C, editors. Clinical data management. 1st ed. New York: Wiley; 1993:133–52.

[18] Spriet A, Dupin-Spriet T. Designing case report forms. In: Good practice of clinical drug trials. Coluzzi R, Young J, translators. 2nd ed. Basel: Karger; 1997:64–76.

[19] Clarke PA. Data validation. In: Rondel RK, Varley SA, Webb C, editors. Clinical data management. 1st ed. New York: Wiley; 1993: 189–212.

[20] Jasperse DM, Ahmed SW. The Mid-Atlantic Oncology Program's comparison of two data collection methods. Control Clin Trials 1989; 10:282–9.

[21] Goldstein JL, Hobbs HH, Brown MS. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Valle D, Sly WS, Childs Kinzler KW, Vogelstein B, editors. The metabolic and molecular bases of inherited disease. 8th ed. New York: McGraw-Hill; 2001:2863–913.

[22] Jansen ACM, Van Aalst-Cohen ES, Tanck MW, Trip MD, Lansberg PJ, Liem AH, Roeters van Lennep HWO, Sijbrands EJG, Kastelein JJP. The contribution of classical risk factors to cardiovascular disease in familial hypeercholesterolemia. J Intern Med 2004;256:482–90.