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.

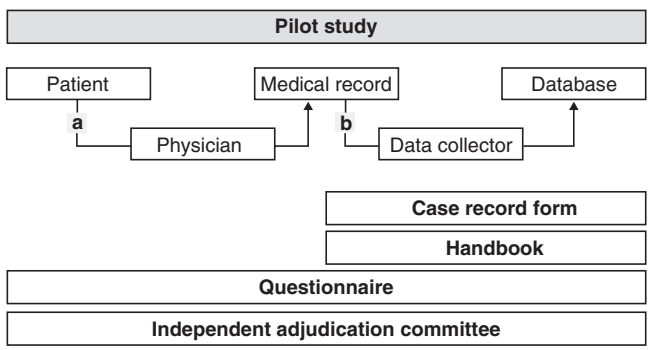

Fig. 1. Data information flows (a) from patient to medical record and (b) from medical record to database.