Large Administrative Database Analysis

Background Paper 2

SUMMARY

Large health administrative databases are used in three different ways to assess the effectiveness of medical treatments: in descriptive studies, in comparative studies, and as adjuncts to other research methods.

In descriptive studies, administrative databases can be used to provide estimates of the rates at which medical treatments are used. The degree to which these rates vary across population subgroups, time periods, and geographic areas can be contrasted. Administrative databases can also be used to provide general assessments of important clinical and economic outcomes experienced by individuals who receive the treatments. Such assessments can sometimes provide surprising results that raise questions about how medical treatments-even well-established treatments-are used.

Some researchers have also used administrative databases in comparative studies, to identiy populations that receive competing types of medical treatment. The populations' health outcomes -e.g., rates of mortality, rehospitalization, or reoperatio-are then compared. These comparative studies, howevever are rarely sufficient themselves to draw definitive conclusions about relative effectiveness, because like other nonrandomized studies their results are susceptible to unrecognized underlying biases that can render the conclusions invalid. Moreovever the quality and quantity of data in existing databases often limit the researchers' ability to use the adjustment techniques employed in other observational research.

This technique is most likely to produce valid results if the medical condition al issue and associated risk factors have been well-

by

Jeff Whittle

Veterans Administration Medical Center

Pittsburgh, PA

studied, if the treatment is applied in a standardized way, and if the data needs are defined prospectively, so that the database is assembled with the research question in mind.

Although administrative databases are severely limited in their ability to be the basis for valid comparisons between technologies, they may substantially increase the weight of evidence about a treatment. They are also useful as adjuncts to other research methods. Using administrative databases as sampling frames, for example, allows researchers to identifiy populations of particular interest for further study. The ability to retrieve data from the medical records of all members of a representative population in a database, or to contact the individuals directly, could be extremely valuable, although it raises privacy issues that must not be dismissed. Another approach entails using a claims database to enhance the followup for a population carefully characterized by a study that entailed primary data collection.

Linking administrative databases with other medical information sources (e.g., cancer registry data), augmenting administrative data with additional elements (e.g., information on health status and functioning), and making other improvements in the availability and accuracy of data could also expand the usefulness of this tool.

One of the earliest demonstrations of the potential power of using routinely collected data on health services was an analysis published in 1938 by Glover (58), who described how the rates at which tonsillectomies were performed on British schoolchildren varied among school districts. Over the course of the past two or three decades, the prominence and volume of such analyses have soared (27), following several developments that facilitated the use of data from claims and discharge abstracts.

Advances in computer systems have allowed large numbers of records to be manipulated at reasonable cost, which has resulted in computers being used to store huge amounts of fiscal and administrative data and has facilitated the development of large registries of diseases and procedures. At the same time, processing these large databases has become less time consuming and expensive, making them more accessible for health services research.

In addition, new computer software (167) allows analysts to use sophisticated statistical techniques. For data sets as large as those now being examined, some statistical tests—such as Cox's proportional hazard analysis of survival data (1 15) and multiple logistic regression (200)—would have been nearly impossible to perform with manual techniques.

Another significant development occurred in 1965, when the United States established Medicare, the nationwide health insurance plan for individuals 65 years of age and older[1] (87,157). Because the vast majority of Americans in this age group are eligible and choose to participate, a sample of Medicare beneficiaries approximates a sample from the U.S. population aged 65 and over (47,65,1 14). Medicare gathers data on most of the health care that is provided to beneficiaries, and Social Security data on mortality can be used to ascertain a beneficiary's vital status after treatment (9). Because each Medicare beneficiary has a unique identification number, his or her use of health services over time can be traced (with some limitations) (114).

The interest in database analysis for assessing medical care burgeoned in response to several factors. First, the aggregate costs of medical care in the United States continued to climb (98). Second, demonstrations of wide variations in the use of common treatments for common conditions

[1] Some of the first claims-based technology assessment research was conducted with data from the Canadian health care system. Canada's national health insurance was introduced in various provinces between 196 I and 1971.

(118,195) suggested that traditional research methods had not defined the best courses of action for many common clinical problems (34,191). Third, when the outcomes of some procedures differed in practice from what had been suggested by the medical literature, analysts recognized the need for more research into how well medical technologies perform in real-world medical practice (175,198,204), as opposed to more limited settings (e.g., academic medical centers). And fourth, the federal government, primarily through the Agency for Health Care Policy and Research (AHCPR), significantly increased funding for this type of research.

AHCPR was established in 1989 to "enhance the quality, appropriateness, and effectiveness of health care services through a broad program of scientific research and dissemination" (184). Prominent among its sponsored research activities are the Patient Outcomes Research Teams (PORTS). These multidisciplinary, multiyear research teams each focus their studies on a particular health condition, and large administrative database analysis has been one of the research tools emphasized by the PORTS.

In addition to being a prime funding source for research into the effectiveness of medical care, AHCPR has encouraged database analysis by facilitating professional communications about its limitations and potential, by carrying out assessments of the existing databases, and by developing databases for use by researchers (184,185). A major function of one branch of AHCPR, the Office of Science and Data Development, is the development of databases as research tools (184).

Types of Databases

This paper focuses on four types of health care databases:

- Claims databases, which consist of claims to third-party payers for reimbursement for medical services provided to covered individuals. The claims can be made for prescription drugs, hospital care, outpatient care, medical equipment, and so on. Claims databases are maintained by third-party payers.
- Discharge abstract databases, which compile summaries of information regarding hospital stays. Each abstract generally includes information regarding the patient's age, sex, and race, the conditions treated during the hospitalization, the procedures performed, and other aspects of the hospital stay, such as the dates of admission and discharge. Discharge abstract databases generally contain information similar to that submitted in claims to third-party payers for reimbursement of hospitalization expenses, although additional data elements may also be included.

  Disease and procedure registries, which contain data regarding individuals who have specific diseases or undergo specific procedures. Disease and procedure registries include all the individuals in a defined population who have the disease of interest or undergo the procedure of interest.
- Practice databases, which contain data accumulated in the course of providing clinical care to patients. All patients receiving care in a particular setting are included, regardless of their diseases or the procedures they undergo. Generally, a practice database consists of a room full of patients' medical records.

Table 2-1 describes these and some other types of large databases that are of potential use for assessing medical care, including some addressed in this paper chiefly with regard to their use with claims and discharge abstract databases. The latter include databases generated as part of large epidemiologic studies such as the Framingham Heart Study (150). (Population surveys, such as those conducted by the National Center for Health Statistics (187), also have a role in medical evaluation but are not discussed in any detail here.)

Role in Evaluating Medical Technologies

The analysis of large health administrative databases has three related but distinct potential applications in efforts to evaluate medical technologies and services.

TABLE 2-1: Some Types of Databases That Could Be Used for Evaluating Medical Technologies

Database type	Description of population included	Data elements typically available	Examples
Claims for insurance payment	All individuals covered by an insurance plan.	Provider, service provided (e.g., a procedure code), reason for service (e.g., diagnosis), charge, payment, patient demographics, patient identifier.	Medicare, Medicaid, private insurance claims databases.
Discharges abstract registries	A defined set of hospital admissions-e. g., all those occurring in a state or all those in hospitals participating in a voluntary registry	Descriptions of hospitalization (including patient characteristics, discharge status, procedures performed, admission and discharge dates), hospital identifier. Certain registries may collect additional data elements (e.g., in New York State, detailed data on catheterization results for patients undergoing CABG.	State discharge registry, VA Patient Treatment File, CPHA database.
Disease registries	All people with certain disease(s) who meet specific criteria (e.g., seen at a participating hospital, resident of a geographic area).	Detailed disease-specific information, patient demographics, patient vital status. May include information on initial treatment. Since these are often gathered at a single site, care received offsite may be poorly recorded; for example, outpatient chemotherapy maybe missed by a hospital-based tumor registry.	Cancer registries, communicable disease reporting systems.
Procedure registries	All people undergoing certain procedures who meet specific criteria (e.g., have procedure done by a provider that is participating in the registry). Thus, these registries are typically not population-based, since not all providers in a region are participating.	Details of procedure (e.g., results of cardiac catheterization, complications of procedure), demographics, vital status in followup Since focus of registry is on the procedure, very detailed data relating to the procedure may be available.	CASS cardiac catheterizatlon registry.
Databases gathered as part of a separate research project	The population identified for study in the original research project.	Patient characteristics, data collected for the original project (frequently quite detailed in the original area of interest, especially if the population included is relatively small), followup data regarding study endpoints. Patient identifiers may not be available because they are destroyed to preserve patient confidentially. Thus, added data regarding the cohort may be difficult to obtain.	Framingham Heart Study, Multiple Risk Factor Intervention Trial screening cohort.
Practice databases	All patients in a given practice setting.	Data gathered in the course of practice (e.g., laboratory tests, physical exam results, diagnoses), demographics, charges payments. These data typically do not conform to a predefine set of data that is gathered on each patient. This disadvantage is weighed against the fact that all of the data available to the clinicians managing the patient are available to the researcher using the database.	Traditional medical records, Duke Database for Cardiovascular Diseases.

KEY CABG = coronary artery bypass graft surgery; CASS = coronary artery surgery study; CPHA = commission on professional and hospital activities; VA = Veterans Administration

SOURCE: Jeff Whittle, 1994.

First, large databases have come to be a staple tool for descriptive *studies* of medical behavior and clinical practice. These include research to describe the variation across areas or populations in the rates with which procedures are done, studies that describe the outcomes associated with a particular procedure or practice, and studies whose goal is to describe the current state of clinical practice. This use of large database analysis is well-established, is used widely in research associated with the federal government medical effectiveness initiative, and is relatively noncontroversial.

Second, large databases have been used to conduct *comparative studies:* studies in which the outcomes of two or more interventions are compared in an attempt to determine which is the most effective. This application has also been promoted under the federal medical effectiveness initiative. However, it is much more controversial than descriptive studies, because of the difficulties in conducting valid comparative studies using observational rather than experimental designs.

Third, large health administrative databases have been used as *adjunct methods* to enhance other research techniques. This set of applications is potentially wide and is only beginning to be explored.

Each of these applications is described below in more detail, along with a discussion of some of the issues and caveats each entails. Many of these issues, such as problems of incomplete or incorrect coding, have been debated and investigated primarily in the context of the use of these databases in descriptive studies. They are discussed here in that context, although they often apply to other applications as well. Other issues, however, are unique to a particular application. This is especially true of comparative uses of large administrative database analysis. Because this use has featured prominently in many of the research projects sponsored by AHCPR, it is discussed in detail below.

DESCRIPTIVE STUDIES

Applications

Variations in Clinical Practice

One important step in assessing medical care is to determine who receives it. Analyzing databases allows researchers to describe the population of patients and to contrast the rates at which subgroups defined by geography, race, sex, or other characteristics undergo particular procedures. The rate at which the use of a treatment changes over time can also be of interest, especially when the increased use of one treatment may be related to the decreased use of an alternative treatment.

Geographic variation

Many demonstrations of variation using claims and administrative databases have been published since Glover reported the variation in tonsillectomy rates across different areas of England (58). Research combining discharge abstract databases with other databases has shown that geographic variation cannot be satisfactorily explained by differences in population characteristics, availability of services, or other structural factors (23,106,130, 158,192,197) (although there is a relationship between the number of providers and the amount of service provided (1 18,194)).

Wennberg has hypothesized that the variation reflects a lack of professional consensus about when treatments are appropriate (19). Even after joint discussions and reviews of the literature, expert clinicians have widely varying opinions regarding whether certain clinical scenarios are appropriate indications for a variety of procedures (13,142). The uncertainty implied by the unexplained variations in the provision of treatment (21 ,193) was a major impetus for the Health Care Financing Administration's (HCFA'S) effectiveness initiative (162) and the formation of AHCPR (97).

The demonstrations of variation have affected practice. After practitioners learned that hysterectomy rates were highly variable among regions in Saskatchewan, Canada, the number of hysterectomies performed by high users decreased (33). Similar results occurred with tonsillectomies in Vermont (190,195). Orthopedic surgeons in Maine are examining indications for a number of orthopedic procedures, whose use has been shown to vary geographically (103).

Variations among specific populations

Claims and discharge abstract databases have also been used to study variations in the provision of certain treatments to population subgroups, including elderly and poor people (201,209), racial minorities (51,62,200), and residents of rural areas (156). Many of these studies have found that some segments of the population are treated less frequently than others, which impels researchers and society to consider the reasons for the variation (72,202).

Although the variations seen in database studies often seem to be important, a precise understanding of the roots of the variations cannot be gleaned from the current databases. For example, the relatively low rates at which invasive procedures are performed on African Americans, compared with whites, could result from differences in coronary anatomy, baseline comorbidity, or patients' preferences—information attainable in prospective studies but not in current databases (162,200,202).

Variation over time

Claims databases have also been analyzed to detect temporal changes in the provision of treatment. One study, for example, showed that the performance of radical prostatectomy had increased nearly sixfold between 1984 and 1990. This implies a significant change in how the procedure is used to treat prostate cancer—a change that is undergoing further evaluation (1 22). Database analyses of changes in the provision of medical care have been used to assess the compliance with consensus recommendations (39,140,169), the introduction of new treatments (155), and the effects of Medicare's change to case-based prospective payment as a means of paying for hospital care (100,172). By demonstrating poor adherence and weak response to the recommendations of the National Institutes of Health (NIH), database analyses have contributed to the growing recognition of the need for research into the dissemination of new technology and information.

Technology substitution

A special type of variation occurs when the use of one procedure decreases the use of another-a phenomenon that is likely to be of particular interest when one of the procedures is new. Unfortunately, very new procedures are often hard to detect in administrative databases. In a recent study of substitution of angioplasty for bypass surgery in the treatment of peripheral vascular disease of the leg, for example, no specific code existed for angioplasty of the arteries to the legs. At a cost of considerable time and money, the researchers had to design and test an algorithm using a combination of diagnosis and procedure codes to identify the patients who had received angioplasties (79).

Analyses of claims and discharge abstract databases can address the question of whether increases in the use of a procedure were associated with decreases in the use of its alternative in a particular population, but not whether the new procedure caused the decrease. In the study noted above, the researchers found no decrease in the rate of peripheral artery bypass surgery as the rate of angioplasty increased overtime, but they could not determine whether the rates of surgery would have been greater had angioplasty not been available. Thus, the actual question—whether the use of angioplasty reduces the need for surgery—was not directly answered. Certain data available in prospective studies but not in the database (information regarding angiograms, clinical conditions, and the like) would have helped researchers answer the actual question.

Outcomes Assessment

Another aspect of evaluating medical technology entails determining the effects of putting a proce-

dure into practice. These effects include both economic and clinical outcomes, whether good (pain relief or improved functioning) or bad (rehospitalization, complications, or deaths). Because the outcomes are likely to differ for patients with different characteristics, an ideal assessment would describe all the relevant outcomes and explain how they vary among patients defined by such characteristics as age, sex, clinical condition, and the setting in which they were treated. These data might then be used to identify groups of patients for whom the treatment's effects were good or were bad.

Databases that can be linked to reliable sources of information about death provide a powerful means of looking at mortality in a defined population. In addition, the rates of hospitalization, reoperation, and certain complications can be determined. Studies of the outcomes of surgical treatment for benign prostatic hyperplasia (BPH) (161,198,199), for instance, were influential in bringing about a recognition of the need for the evaluation of common procedures (162). Furthermore, a decision analysis combining these data with primary data regarding symptomatic outcomes clarified the importance of patients' preferences in selecting management options (53).

Similarly, reports that short-term morbidity and mortality following carotid endarterectomy were higher than expected in the Medicare population may have contributed to a trend toward lower rates of the treatment nationwide (8 1,205) (and to a decline in the enrollment rates in a randomized trial of the treatment (1 l)). Studies of Medicare patients' outcomes during hospitalization have been another kind of influential (though controversial) assessment.[2]

Issues and Limitations

Coding Issues

Much of the concern about claims database analyses has focused on the coding system used to represent diagnostic and procedural information (40).[3] Hospital discharge data in the United States are coded using the *International Classification of Diseases, 9th Revision-Clinical Modification (ICD-9-CM) (181). The ICD-9-CM* includes more than 10,000 numeric codes, with as many as five digits apiece. All five-digit codes are subsets of four-digit codes, which are subsets of three-digit codes. The three-digit codes were organized into 17 chapters representing broad disease categories, which range from "neoplasms" to "symptoms, signs and ill-defined conditions"4 (40).

The information represented by these codes is the basis for hospital payment by Medicare, many national health statistics, and other uses. The coding system is updated periodically in order to meet reimbursement needs, to allow more precise identification of diseases and procedures that have grown in significance, or to clarify how certain diagnoses or procedures should be coded.[5]

The time lag before the implementation of the coding changes that are needed to identify new diseases or procedures causes problems for database researchers. The acquired immunodeficiency syndrome (AIDS) first received a specific diagnosis code in 1986, eight years after the first case reports of AIDS were published and three years after the etiologic agent had been identified. Similarly, new procedures may be part of practice for some time before new ICD-9-CM codes are devised to describe them specifically. For example, percutaneous transluminal coronary angioplasty (PTCA)

[2] These assessments had been a part of the annual HCFA hospital mortality reports, which were suspended in 1993.

[3] For more detail on this issue, see any of the several reviews that have been published (40,84,175).

[4] For example, the chapter "Diseases of the Digestive System" includes the code 532 for "duodenal ulcer." Code 532.0 refers to an acute, bleeding, duodenal ulcer, and 532.01 refers to an acute, bleeding, duodenal ulcer with obstruction.

[5] An ICD-9-CM Coordination and Maintenance Committee was established in 1985 to advise HCFA and the National Center for Health Statistics about the need for changes in the coding system.

was initially coded to a procedure category that was also used for open-heart surgical procedures. The use of this code caused the patients who underwent PTCA to be assigned to a Medicare payment category that was reimbursed at a level much higher than that of the usual costs (173). New procedure codes created in 1986 placed the procedure in a more specific (and less generously reimbursed) category.

To limit the number of codes in the system, the ICD-9-CM lumps certain entities, which can obscure important differences. Codes for patients with renal dysfunction, for example, distinguish between acute or chronic cases but not among levels of dysfunction, which range from a slight change in a biochemical test that has only minor functional effects to complete cessation of kidney function (35). Moreover, the ICD-9-CM does not systematically include the sidedness (left or right) of a disease or a procedure. Consequently, researchers conducting database studies of procedures that can be done on either side (e.g., cataract surgery) may have difficulty interpreting whether certain procedures or diagnoses that occur after the procedure of interest are related to it (92).

Another limitation in the ICD-9-CM coding system is that different codes can sometimes be used to describe the same condition. For example, a code for a symptom (angina), a disease process (myocardial ischemia), or an anatomic abnormality (coronary atherosclerosis) can all be correctly and legitimately used to describe a patient with narrowing of the coronary arteries that causes chest discomfort with exertion (175).

A limitation of hospital discharge data is that they do not reveal whether coded conditions were present at the time of admission (preexisting conditions) or developed during the hospitalization (possible complications). To address this problem in New York, coders were asked to indicate whether conditions were present at the start of the hospitalization, but initial studies show that the coders have been slow to implement the change (61). An alternative approach is to exclude those conditions that could develop as complications from being considered as comorbidities. Unfortunately, many of the most important factors affecting an individual's baseline condition fall into that category.

In addition to dealing with coding problems in individual databases, researchers analyzing more than one database for a particular study may have to resolve differences in the coding systems. Procedure codes in the ICD-9-CM, for example, do not correspond to the coding system used for most professional service claims[6] (2).

Inaccurate Data

The accuracy of a database's coding depends on how often the codes entered into the database are the codes prescribed by the rules of the coding system. According to studies of coding accuracy that were conducted before Medicare's prospective payment system (PPS) for hospitals was introduced (in 1983), patients' age, sex, admission date, and discharge date were generally accurate, but the diagnosis and procedure codes were not (31,89,90). Even at the three-digit level, more than 25 percent of the principal diagnosis codes were different from those assigned by expert reviewers (90).

Financial incentives for complete coding were introduced with the Medicare PPS, which linked the amount of payment to a patient's diagnosis, and coding accuracy did improve. For example, at the three-digit level, overall agreement on the principal diagnoses increased from 73 percent in 1977 to 78 percent in 1985 (47). Subsequent data suggest that accuracy has continued to improve (79,80).

---

[6] HCFA mandates the use of the Current Procedural Terminology (CPT) coding system for professional service claims and supplemented the CPT with additional codes for nonphysicians' services, which it did not include. This augmented CPT is known as the HCFA Common Procedure Coding System (HCPCS). In addition, local intermediaries (insurers) can create codes with the approval of HCFA. This complicates analyses that cross states, as well as those that combine inpatient and outpatient data.

Accuracy apparently varies significantly among diagnostic and procedure codes. In the 1985 Medicare discharge data, 90 percent of the patients coded as having lung cancer had actually been diagnosed with lung cancer, and only 7 percent of patients who had actually been diagnosed as having lung cancer had not been coded as such. By contrast, peripheral vascular disease was coded for fewer than 60 percent of the patients who had the condition, and only 53 percent of those coded as having the disease actually did (47). Major procedures, which often affect Medicare payment, are quite accurately coded. For example, 96 percent of the patients who underwent coronary artery bypass graft (CABG) surgery were coded, and when the procedure was coded as having been done, it had always been done. Minor procedures, however, are much less reliably recorded. One study identified far fewer individuals coded as receiving total parenteral nutrition (a form of specialized intravenous feeding) than were known (from other data sources) to be receiving the treatment (120).

It is unclear whether improvements in coding that accompanied the implementation of the Medicare PPS are reflected in codes not used for Medicare payment. Although many other databases have been subjected to review (25,1 17), the results are seldom published (101).

Financial incentives can sometimes result in one-sided errors. For example, the Medicare PPS pays more for more severely ill patients, and the study of coding accuracy in 1985 discharge data showed that errors in how severity of illness was coded systematically tended to overstate severity, increasing hospitals' reimbursements (80). (A study of 1988 discharge data did not reveal the same tendency, perhaps because of the strict laws that now require attending physicians to certify the accuracy of the designated diagnoses (79)).

The accuracy of the coding on claims for professional reimbursement may be somewhat higher, because professionals tend to perform the same procedures (and use the same codes) repeatedly. In a recent study of carotid endarterectomy, in which professional claims were used to identify patients (205), codes for the procedure were verified for more than 95 percent of the patients. Diagnostic information is seldom available for professional claims, however, and has very rarely been used in research.

Studies of temporal and geographic variations may be subject to bias because of disparities in the quality of coding overtime or among regions (44). If coding is unusually complete in one area, the rates at which procedures are performed in that area may appear to be unusually high. Both selective coding to maximize reimbursement(171) and the trend to move some procedures from inpatient to outpatient settings (163) could create a false appearance of temporal trends in treatment rates if inpatient databases are used.

In addition to errors in coding the diagnoses that are recorded in patients' charts, physicians themselves sometimes make diagnostic errors, and one doctor's diagnoses are likely to differ to some extent from those of another. The lack of precise definitions leads to wide variations in the reported incidence rates of diseases, depending on the criteria used for making the diagnoses (40). In a discharge abstract database, a correctly coded diagnosis means only that the attending physician made that diagnosis, irrespective of whether the appropriate diagnostic criteria were used.

Differences among physicians in how completely they choose to evaluate their patients can lead to bias, because patients who have more complete evaluations are more likely to be found to have signs that indicate poor prognoses, for example, the spread of cancer from the site of origin. The prognosis for patients whose spread of cancer has been found only after extensive testing falls between the prognosis for patients in whom the spread of cancer is obvious and that for patients whose cancer has not spread. Moving this intermediate group from the good prognosis (no spread) group to the poor prognosis group improves the prognoses of both groups[7] (41).

[7] Alvan Feinstein named this the Will Rogers effect, after the humorist observation that the move of many Oklahomans to California was increasing the average intelligence in both places.

Groups defined by their clinical characteristics or treatments may receive systematically different evaluations. For example, testing to rule out the spread of lung cancer may be more extensive for patients treated with surgery than for patients treated with radiation therapy. If so, a comparison of how the two groups fared could be biased.

Identifying the Relevant Population

Another important issue with which researchers must contend involves being able to identify correctly the universe of people of interest. For example, when dealing with the rates at which treatment is provided, researchers must not only know the denominator (the population under investigation) but also be able to identify the numerator (the people who receive the treatment). The numerator in a claims study is usually defined as those individuals who are coded for the treatment of interest. The appropriate denominator is not always clear. One simple denominator comprises all the people whose receipt of a treatment could appear in the database. Such a denominator usually includes everyone in the population, unless the treatment of interest is only performed on members of one sex. If a procedure can be done only once, however, anyone who has already had the procedure should not be included in the denominator. Thus, if hysterectomy rates are to be described, the denominator should include only women and should exclude any woman who no longer has a uterus (because of prior hysterectomy). The fact that as many as half of some female populations have had hysterectomies might bean important source of variation (158).

Often, a different denominator—people with the medical condition for which the treatment is provided—may be more appropriate. In a study of variation in rates of radical prostatectomy,[8] for example, a useful denominator would be men with prostate cancer (122). Similarly, because CABG surgery is performed only on patients with coronary artery disease, the analysis ideally would cover only such patients (202).

Defining the appropriate denominator can be difficult. Researchers conducting claims-based analyses of variation in the rates of treatment have often assumed that similar proportions of different populations are at risk, perhaps after accounting for differences in the age and sex distributions. This assumption allows the total population, which is easy to quantify, to be used as the denominator, but the assumption is not always valid.

Although some analyses of variation have used patients hospitalized with conditions of interest as the denominator, many similarly afflicted patients are not hospitalized. The interpretation of the denominator, therefore, becomes difficult. Nonetheless, such conditions as myocardial infarction (180), childbirth (119), and hip fracture (45)-which are almost always treated in the hospital if they are recognized-can define more complete denominators. Otherwise, the differences in hospitalization practices across time, regions, or population subgroups could cause spurious apparent variations in the rates.

For many procedures, appropriate databases are hard to find. Outpatient procedures, for example, are not included in statewide discharge databases. HCFA now retains all professional claims under Medicare, including bills for procedures performed in physicians' offices, but this is a recent development, and the data go back only to 1991.

Other procedures that are difficult to assess are those performed on hospitalized patients but not reliably recorded (e.g., total parenteral nutrition). New procedures are especially problematic, because until they have been assigned unique codes, they cannot reliably be identified in claims databases.

Even a well-defined population can receive some care that is missed if a researcher uses only a single database. Some individuals, for example, may be covered by both their own insurance and

[8] Radical prostatectomy is a surgical procedure used to treat localized prostate cancer. It is significantly more dangerous than the procedures used to treat benign prostatic hypertrophy.

that of their spouses. Others may lose their coverage or never have any at all. Restricting a study to the consistently covered portion of the population raises the possibility of selection bias (164). Medicare data are particularly useful because nearly all the recipients continue to be enrolled until they die. Even then, however, the beneficiaries may receive treatment from providers (e.g., the Department of Veterans Affairs (VA)) whose services do not appear in the Medicare database (49).

Obtaining Outcomes Information

Although outcomes assessments have been important products of database analyses, a number of limitations are obvious. In addition to problems with coding and with services provided by sources not covered in the databases, many important outcomes other than death are not included in claims and discharge abstract databases. When such outcomes are the appropriate measures for comparing the benefits of different treatments, database analyses are difficult.

For instance, total joint replacement--considered the most important advance in the management of arthritis in the past 20 years (55)—almost certainly does not increase the survival rates of patients with arthritis and might even decrease the rates slightly (because the disease is not fatal, whereas surgery carries some risk of death) (204). The objective of the procedure is to improve the patient's quality of life by relieving pain and increasing mobility (63). Unfortunately, neither claims nor discharge abstract databases include measures of these outcomes. Among the other important outcomes not available in such databases are the relief of such symptoms as incontinence and diarrhea, the ability to function socially, and a sense of well-being.

Alternative Approaches

There are alternative methods for examining patterns in the use of medical services. One approach, suitable for procedures that require a single expensive piece of equipment (such as an artificial hip joint), is to survey the manufacturers of the equipment. If there are only a few suppliers, relatively good estimates of overall rates of treatment seem possible (74). Another approach is to survey a sample of providers regarding the frequency with which they provide the treatment (32,74). Unlike database studies, these approaches do not require coding conventions to identify the treatment. A third alternative is to sample a population to identify individuals who have been treated with a particular procedure. This allows researchers to collect precisely the variables that are of interest in characterizing both the numerator and the denominator.

These approaches have their own drawbacks, however. Studies of the use of a procedure at a single facility or a few hospitals, for example, might include relatively few patients who have undergone the treatment. The denominator population from which the patients were drawn would be hard to define, and the treatment rates at the participating sites might not be representative. Several of the health surveys conducted by the National Center for Health Statistics (e.g., the National Medical Care Utilization Survey) could be used to study treatment rates, but the surveys are relatively expensive and time-consuming, include relatively few individuals, and provide only limited details about which medical services are used. Primary data collection to address these concerns would probably be prohibitively expensive. Moreover, the data collection would have to be continued if temporal trends were of interest.

COMPARATIVE STUDIES

Frequently, when researchers assess a treatment, the most difficult question to answer is whether the outcomes (mortality, morbidity, cost) for patients treated with therapy A are better than those for similar patients treated with therapy B (or for patients who do not undergo treatment). The use of claims databases to address this question has generated much controversy (7,15,16,66,70,1 10, 124,125. 178).

I Rationale

For researchers attempting to assess the comparative effectiveness of health technologies, database analysis offers a number of potential advantages over other methods. These advantages helped generate the enthusiasm for using database analysis in effectiveness research. Studies comparing the outcomes of transurethral prostatectomy (TURP) and open prostatectomy on men with BPH, for example, stimulated a reexamination of the question of which treatment is the most appropriate (46,73). A recent study of outcomes associated with management of cataract surgery patients showed that the rate of retinal detachment was several times higher among patients who had undergone posterior capsulotomy[9] than among patients who had not undergone the procedure (93)--information that, if confirmed, could enhance decisionmaking regarding the timing of the procedure.

Large Size

Probably the most obvious advantage of using large preexisting databases to conduct comparative effectiveness studies is that the databases are large. More than 25 million patients are represented in the Medicare claims database, and about 10 percent of the U.S. population lives in the areas covered by the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) cancer registries. Analyses of such large databases can provide estimates of the rates of even relatively uncommon events (e.g., the adverse effects of particular drugs (147,176) or the complications that can occur after certain types of surgery (92,154)) or can identify cohorts of patients with rare conditions (e.g., endocarditis (7)). The size of the databases also allows researchers to subdivide groups of patients by age, race, or sex and still have a significant number of subjects whose experiences with treatment can be studied (180).

Representative Samples

Because they are generated routinely in the course of providing care to patients, claims and discharge abstract databases may include everyone with the condition of interest in the populations for which the databases are maintained. For example, Medicare claims data cover more than 95 percent of Americans over the age of 64 (1 14). State discharge abstract databases generally cover nearly all the hospitalizations that occur in the state. The people or hospitalizations covered in such databases are generally much more representative than the populations studied at individual facilities or at a few academic medical centers (175). The factors that cause patients to enter tertiary medical centers for care are often related to the results of the treatment they undergo (164). Thus, studies using Medicare claims data to determine the short-term mortality rates following pneumonectomy for lung cancer (203), carotid endarterectomy to prevent stroke (204), and transurethral prostatectomy (TURP) for BPH (161) have found higher mortality rates than those found in studies of patients who were treated at medical centers that had particular interests in the diseases (50,57,132).

Opportunity for Followup

"Long-Term Follow-Up is a Problem" is the title of a 1983 editorial in the *American Journal of Public Health (5).* This statement holds true for a number of outcomes of great interest in assessing medical care. Mortality, cost of care, health status, and rehospitalization long after treatment are important in defining a treatment's utility. Because individuals who are not followed up may differ systematically from the rest of the population being studied, a major portion of research efforts are directed toward assuring complete followup (54). The concern is so great that most investigators try to exclude patients who are unlikely to follow up

[9] Posterior capsulotomy is frequently done following extracapsular cataract extraction as a treatment for posterior capsular opacity, a common complication following extracapsular cataract extraction. Since posterior capsular opacity can vary in severity, the decision to perform posterior capsulotomy relies in part on a subjective assessment of the opacity's effect on vision.

reliably, despite the possibility that their exclusion could affect the representativeness of the population (83).

In some cases, insurance databases can provide more complete followup at considerable y less cost. For example, the Medicare claims database includes data, gathered by the. Social Security Administration (SSA), regarding the vital status of beneficiaries (199). Because vital status is important for determining Social Security payments, considerable care is given to ascertaining whether and when beneficiaries have died. The expense of this ascertainment is borne by the SSA, which frees researchers from a heavy burden.

Less "Attention Bias"

Many aspects of medical care are difficult to study, because physicians and patients may change their behavior if they know that they are being studied. Physicians may practice more cost-effective medicine, schedule more timely follow-up visits, or provide more preventive services. Patients may be more likely to stop smoking, adhere to a recommended diet, or comply with a complex medical regimen (83). This has been called "attention bias" (164) or the "Hawthorne effect" (after the Chicago industrial site where research in the 1920s showed that productivity improved when workers were being observed) (151).

By using data gathered in the course of routine clinical practice, researchers can avoid this phenomenon. The providers and patients involved do not know that they are the subjects of a study. Indeed, at the time the care was delivered, the patients were not in a study, inasmuch as the study began sometime later.

Timeliness

Because claims data are needed immediately for payment purposes, they become available for research fairly quickly. Thus, Medicare claims data from 1991 and VA data from fiscal year 1992 were available for research in 1993. Because data regarding several years' worth of patient followup are included in the Medicare and VA databases, researchers can relatively easily conceive of and carry out studies to address questions that have been newly recognized, even if they involve events that have occurred long after the treatment (e.g., long-term survival following lung cancer resection). In contrast, a study using primary data collection may have to wait many years for a sufficient number of patients to be identified and to experience the outcomes of interest.

Another way in which database analyses can be timely derives from their power. Treatments can change rapidly, particularly when they are new. By the time a randomized controlled trial (RCT) of a treatment is complete, the treatment under evaluation may not be acceptably close to the treatment that has become state-of-the-art since the RCT began. The results of coronary artery bypass graft surgery, for example, changed significantly as the technology evolved. The first major randomized trial of CABG found excessively high rates of surgical mortality, but the results of that study do not seem to apply to surgery done with modem techniques (146). By contrast, the numbers of patients covered by databases are so great that researchers can use data from a short period when the technology is likely to be relatively stable. Thus, database analysis can address the moving-target problem that plagues other approaches to assessing health care.

Validity and Reliability

The relative ease of database analysis makes it an increasingly attractive method to address clinical questions, particularly when the outcomes of interest are rare but are likely to result in events that the databases can reveal (92). The important question, however, is whether database analysis can provide valid answers to these questions (1 5).

Using Observational Data for Comparative Studies

Differences in the outcomes of alternative treatments do not necessarily mean that one approach is superior, unless the way the treatments are provided and the composition of the populations receiving them are comparable. A sick population provided with a superior treatment might well fare

worse than a healthy population provided with an inferior treatment.

Much scientific discussion has focused on the validity of comparing the outcomes experienced by apparently similar groups of patients. For example, researchers often compare the results for patients receiving a new treatment at a particular center with those for historical controls-patients who had the same condition and received the standard treatment at the same center in the past (56,165). The outcomes experienced by a series of patients treated at a center can also be compared with the outcomes experienced by similar patients as reported in the literature or by contemporaneous patients who have undergone the alternative therapy.

In each of these examples, the use of the new treatment is not the only thing that truly differs for the groups being compared. Historical controls were diagnosed with older tests, which might not have discovered their disease until it had reached more advanced stages. The historical controls also received older supportive therapy, which means that improved outcomes might simply reflect the general improvement in medical care over time, rather than derive from the treatment being evaluated. Similar concerns exist about controls selected from the literature (they reflect the experience at different centers, often from earlier times) and contemporaneous controls from the same institution (the decision to use a new treatment may by influenced by the severity of the patient's condition).

Experimental studies—specifically, RCTs—address these problems of comparability by ensuring that the treatments are performed on groups of patients who are only as different as the random play of chance would allow. This approach permits any eventual differences in outcomes to be assessed by the question of how likely it is that the set of outcomes would have occurred by chance alone.

Although current opinion holds that an RCT provides the most scientifically rigorous way to compare treatments (8,1 12), it has a number of disadvantages. These include the practical (177) and ethical (67,144) difficulties of enrolling patients when there are strong theoretical reasons or clinical suspicions that one treatment is better than another (56), the long delay before the results are available, and the limited numbers of patients often included. Moreover, such studies can be expensive to carry out due to the need for screening many candidates to identify a few eligible patients, the need for extensive quality control activity to make sure that the interventions are applied in a standard fashion, and the practice of obtaining extensive data regarding each patient.

Just as the study's design ensures that the participants are relatively homogeneous, it also limits the researchers' ability to generalize an RCT'S results to patients who do not meet the eligibility criteria. When the eligibility criteria do permit participation by identifiably different subjects (both men and women, for instance, or patients whose disease differs in severity), the relatively small number of patients often makes it difficult to determine whether the results apply equally to all the subgroups.

As an example, a recent trial of lowering cholesterol with cholestyramine showed that the treatment reduced rates of myocardial infarction among 3,806 men between the ages of 35 and 59, the vast majority of whom were white (150). A similar trial in an elderly, female, or African American population would be hard to justify—both on ethical and economic grounds—although some researchers have questioned whether the trials can be extrapolated to populations that were not studied (151). These limitations have spurred interest in designing RCTS that keep the benefits of randomization while incorporating some of the generalizability of database research (1 7). (See J. Burning, M. Jonas, and C. Hennekens, "Large and Simple Randomized Trials," background paper no. 3 in this volume).

When RCTS have not been carried out, what role should nonexperimental data play in the assessments? Should the data be used only to extend the results of RCTS to population subgroups not included in the original trials? Should observational studies never be used to make comparative judgments? Or can fair comparisons of treatments be based on the outcomes experienced by popula-

tions "assigned" to different treatments in nonrandom ways?

The problems with using observational data in general for comparisons offer insights into the problems with using claims databases—a particular source of observational data—for comparative analyses. The most weighty concern about using observational data to compare therapies is summarized by Byar: "In medicine, the doctor chooses the therapy precisely in order to affect outcomes" (123). The choices generally differ for different patients. One way to address the problem is to carefully note all the important differences between the groups receiving the treatments being compared and then to use statistical techniques to correct for the differences. This presupposes that the important differences can be identified and are recorded when the data are collected. Researchers using historical controls, literature controls, or contemporaneous controls have taken this approach when comparing treatments.

At least three more potential problems with comparisons based on observational data have been noted. First, the ability to characterize the treatments being compared is limited, because providers may vary in the way they apply the same treatment, whereas both the new and the control treatments are administered in standard fashions in traditional RCTS.

Second, patients are assigned to one group or another at the time of randomization in RCTS, but group assignments may be ambiguous in observational studies. For example, if a patient is treated medically for coronary artery disease and then undergoes surgery because the medical therapy was ineffective, the patient's subsequent death might be attributed either to the medical group or to the surgical group. Approaches to this problem have been investigated by researchers using the Coronary Artery Surgery Study registry of patients undergoing catheterization for possible coronary artery disease (20), but no course is completely satisfactory. One feasible approach in a prospective study would be to ask the physician to outline a plan of action at the time followup begins. Researchers would then use the plan in assigning the patient to a treatment group.

Third, the ability to characterize subsets is limited by the data that are collected: if new, important risk factors are discovered, they cannot be presumed to be present to the same degree in the two comparison groups. Similarly, the level of detail in which the data are obtained—both at the baseline assessment and in the followup-is likely to be inconsistent unless the investigator prescribes, in advance, what data are to be collected. Worse yet, the level of detail is likely to be inconsistent in a nonrandom way. It is quite plausible that more tests or more followup visits, through which complications could be discovered, would be ordered for patients receiving the new treatment. Whether differences in the baseline characteristics result from biased assessments is difficult to ascertain.

Randomized vs. Nonrandomized Studies

One way to gauge whether valid comparisons of treatments can be made without randomization is to look at instances of comparisons made with nonrandomized and randomized study designs and try to draw conclusions about whether they are equally valid.

There is considerable agreement that spectacular effects (137) do not require randomized assessment. Most of the major advances in cancer chemotherapy, for example, were made without the benefit of randomized trials (56). Similarly, the treatment of endocarditis and tuberculous meningitis with antibiotics could be recognized as major advances without RCTS (70,133). What these cases have in common is that before the availability of the treatment, the patients uniformly fared badly and that the treatment considerably improved their chances of survival.

This does not mean that the comparisons were fair. In fact, people diagnosed with Hodgkin's disease today almost certainly have better prognoses, on average, than did individuals diagnosed with the disease before the advent of chemotherapy, but the availability of chemotherapy is not the sole reason for the improvement. Modem imaging techniques enable the diagnosis to be made earlier, the treatment of infectious complications has im-

proved, and patients are more likely to seek medical care earlier in the course of the disease.

Nonrandomized comparisons have shown treatments to be beneficial, only to have subsequent trials demonstrate them to be useless. Gastric freezing for peptic ulcer disease (13S) and internal mammary artery ligation for coronary artery disease (24,59) are frequently cited examples. Recent studies by careful investigators have revealed similar patterns of unreliability as well. Major trials involving patients with advanced cases of non-Hodgkin's lymphoma have concluded that a first-generation chemotherapy regimen, CHOP[10] (19), is as good as, or better than, more complex and toxic second- and third-generation regimens (48,60). Previously, the results of the newer regimens were compared with the results of CHOP reported in the literature. These reports, and accompanying editorials, had strongly implied that the newer regimens were superior (26,105,170). Despite an acknowledged need for randomized trials, practitioners began using the more complex regimens. A 10-year lag occurred between the early reports and the publication of the studies showing that the supposed benefits of the more complex regimens were not real (48,60,121).[11]

Another modem example of unreliable observational data is particularly interesting because of the methodologic care that was taken to avoid any identifiable biases. Researchers evaluated whether administering lidocaine prophylactically to patients with acute myocardial infarction helped prevent arrhythmia deaths (76). The observational study used stringent entry criteria, a well-defined endpoint, and an adjustment for the differences in the risks associated with the endpoint. The data were collected by trained researchers who were blinded to the study's hypothesis. Nonetheless, the finding that lidocaine had a beneficial effect was not borne out by subsequent RCTS or a meta-analysis of all the available RCTS (68,207). Although neither the RCTS nor the meta-amdysis had sufficient statistical power to identify arrhythmia deaths, the standard interpretation of the available data has been that lidocaine usually should not be administered prophylactically in the treatment of acute myocardial infarction (10).

On the other hand, in several instances nonrandomized comparisons have yielded results similar to those of randomized comparisons. A study that addressed the use of tonsillectomies for children with recurrent sore throats (141) provides an especially good example because the random and nonrandom comparisons were carried out simultaneously at a single institution. When the parents of children who were eligible for the randomized comparison decided against randomization, the children received the therapy their parents requested. The initial evaluations and data collection processes, however, were identical for both the randomized and nonrandomized patients. Moreover, the two groups were followed up in the same manner, including the frequency of visits, the definitions of the endpoints, and the management of throat infections. The researchers compared subgroups matched for important predictors, such as age and frequency of episodes in the preceding two years, and found that the outcomes experienced by the randomized and nonrandomized patients were indistinguishable.

Another study, which used the Duke Database for Cardiovascular Disease (71), was explicitly designed to test the hypothesis that observational data could be used to make fair comparisons between groups assigned to different treatments in a nonrandom fashion. The treatments compared were CABG and medical management for patients with coronary artery disease. These treat-

[10] CHOP, one of the first combination chemotherapy regimens introduced for lymphoma in the mid-1970s, includes cyclophosphamide, vincristine, doxorubicin, and prednisone.

[11] It should be acknowledged that many oncologists do not believe the results of the RCT comparisons. Even with RCTs, purportedly definitive results do not always prove definitive.

ments had been compared in three separate RCTS, whose results could be compared with those of observational studies.

The researchers used the Duke database to identify individuals who would have been eligible for each of the RCTS. The predicted survival rates of these patients, first assuming that all of them had received medical therapy and then assuming that surgery had been performed on them, were calculated using a statistical model derived from the overall Duke database. The survival curves were then compared with the actual survival rates of the participants in each arm of the trials. Differences between the Duke database analysis and each of the RCTS were within the limits of random variation. In addition, the five-year mortality rate of nearly every subgroup of the Duke patients with varying severities of the disease, whether managed medically or surgically, differed from the rates in the RCTS by no more than would have been expected by chance alone.

Still another study assessed the use of beta-blockers after heart attacks, an intervention that has been shown to be beneficial in a number of RCTS (77). The observational study was modeled directly on a specific RCT, the Beta-Blocker Heart Attack Trial (BHAT) (6). Patients in the observational study were patients consecutively admitted to the Yale-New Haven hospital for acute myocardial infarction between 1978 and 1982, while the BHAT enrolled patients from June 1978 to October 1980. The authors of the observational study explicitly compared their results to those of the BHAT (77) and found no significant differences between them, after adjusting for age and severity of disease in the groups who received beta-blockers and those who did not. Each study showed reductions in both 24- and 36-month mortality rates. Moreover, the magnitude of the difference was very similar in each trial.

Lessons

The fact that fair comparisons were made using observational data does not guarantee that similar designs would lead to fair comparisons in other studies. Without RCTS for comparison, how can we determine whether observational data are reliable?

The examples provide some clues. The tonsillectomy study (141) points out the need for good followup, particularly when the outcome of interest is not death. In claims and discharge abstract databases, outcomes other than death are detectable only if the patients or physicians have taken some action. An outcome like “need for repeat surgery” may be biased by a greater level of surveillance for one of the study groups (164). The methods for determining outcomes should be identical for groups receiving alternative treatments. The tonsillectomy study also had the advantage of patients with few comorbidities, and the risk factors thought to predict the outcomes of interest were precisely determined at the outset of the trial. The clinicians involved had little bias toward one treatment or the other, and parents were advised that the randomized trial was an appropriate option. Therefore, there is little chance that the clinicians would have encouraged patients with better prognoses to opt for a particular treatment.

Although the salient risk factors for death from coronary artery disease do not predict even half of the variability in who lives and who dies, they are well studied (69,70,7 1). Clinicians making decisions are unlikely to consider any factors that do not appear in the model used by the researchers who analyzed the Duke database. In other words, the other factors that predict outcomes are unlikely to vary among groups assigned to different treatments because no one knows what those factors are. Their distribution should be random. It is difficult to disprove the contention that clinicians can detect and interpret subtle differences among patients that cannot be captured as concrete data and incorporated in a model, but the Duke researchers have shown that their risk-adjustment model is better than expert clinicians at predicting the prognoses for patients with coronary artery disease (107,1 16). It would be unreasonable to think that these same clinicians could select the patients with the best prognoses for the group that is to receive CABG.

The researchers in the beta-blocker study emphasized several aspects of their method (77). The eligibility criteria in their observational study, and particularly their definition of myocardial infarction, were identical to those used in the BHAT. The researchers excluded any patient who would have been excluded from the BHAT, and they carefully considered how to assign zero time—the time at which each patient's baseline was established. After zero time, followup events were assigned according to their treatment groups. The researchers emphasized that this selection should duplicate, as much as possible, the time at which the random assignment would occur in an RCT. As in the BHAT, adjustments were made for differences in known confounding variables, such as age and baseline prognoses.

The studies in these examples have several important features in common:

1. Each of the clinical entities (tonsillitis, coronary artery disease, myocardial infarction) is well studied and understood.
2. The treatment under study was likely to be applied in a fairly standard fashion. The control group was also likely to receive state-of-the art management, because the control patients and the experimental patients were treated in the same sophisticated settings (although this does not guarantee standard care).
3. The data needed to define the baseline prognoses were collected prospectively and purposefully for each of the observational cohorts. Thus, from the beginning of the data-collection process, the database was oriented toward the type of study that was eventually performed.

Despite the fact that the study of lidocaine incorporated many of these positive features, the results differed from those of the RCTS. The least hopeful interpretation is that nonrandomized studies sometimes provide the right answers and sometimes provide the wrong answers, and that there is no way to tell the difference without an RCT to determine the true answer (15,16).

This interpretation suggests that the usefulness of databases is determined by the answer to a simple question: "Can valid comparisons be made with observational data alone?" Phrasing the question this way, however, ignores the fact that such comparisons are based on careful assessments of patterns of data. In general, the results of a single RCT are not definitive. Indeed, the validity of the comparison of CHOP with the advanced chemotherapy regimens has been challenged by a number of respected experts (121). Thus, many authors would argue that rigorous comparisons using observational techniques have a role in assessing medical treatments, because they can contribute data to the pattern, even if the results of observational analyses alone are not definitive (76,77).

Several differences between the data in these observational studies and the data in claims and discharge abstract databases bear emphasis. Each of the researchers used careful, quality- controlled methods of collecting data elements that had been defined prospectively as appropriate for the problem at hand, but database researchers must use whatever data are available. In general, even databases supplemented with clinical data have difficulty yielding answers to questions that were not formulated carefully before the data were collected (69).

For example, using a dataset that included detailed clinical data from the MedisGroups medical severity classification system in addition to routine hospital-discharge data, researchers examined how patients whose coronary artery disease was treated with percutaneous transluminal coronary angioplasty[12] fared in comparison with patients who were treated with CABG surgery (64). After being adjusted for differences between the groups in the age of patients and the presence of a number of clinical variables, the data showed

[12] Angioplasty is the procedure in which a balloon is inflated in an artery that has been narrowed or closed by a pathologic process (usually atherosclerosis). It usually decreases the degree of narrowing. PTCA is angioplasty of the arteries to the heart.

that the patients who were at low risk of dying did better with PTCA than with CABG. Despite the size and detail of this database, however, the lack of prospective collection of important data relevant to outcomes (e.g., the number of coronary arteries that were blocked) makes it difficult to draw conclusions.

The current claims databases' limitations for making comparisons have been well documented. There is evidence that discharge data restricted to ICD-9-CM codes may not contain adequate detail to allow for valid comparisons of outcomes across treatments (28) or hospitals (111). For example, the five-year survival rates were lower for men whose BPH was treated with TURP than for men treated with open prostate surgery, and an adjustment based on ICD-9-CM-coded discharge data did not explain the difference. This suggested that the open procedure was superior, but an adjustment using more detailed information obtained by reviewing charts showed that the differences probably derived from the higher numbers of important comorbid conditions in the population undergoing TURP (28). Another study found that case-mix adjustment using ICD-9-CM codes explained much less of the variation in hospital mortality than did case-mix adjustment using additional data abstracted from the clinical records (11 1).

Other studies have suggested that using ICD-9-CM codes in making adjustments for differences in groups might sometimes actually produce misleading results. Researchers found that the presence of a number of comorbid conditions actually improved the survival rates of patients admitted for several serious illnesses (95). The researchers interpreted their results to mean that certain diagnoses were usually coded only when there were no more important diagnoses to be coded. At the time of that study, the discharge abstract contained only a limited number of fields in which diagnostic codes could be recorded, but a later study found that the same thing occurred when the number of available fields was not restrictive (85).

Comparisons after adjustments for differences in baseline risks of poor outcomes are almost certainly more valid than comparisons made without such adjustments. Improved methodologies, including hierarchical modeling and instrumental variables techniques, increase the believability of the adjustments[13] (129). Comparisons are also more believable when the medical condition under study is well understood, when the variables are objectively defined, and when the data collection is complete and accurate. Unfortunately, most claims and discharge abstract databases do not currently meet that standard. Moreover, enhancements that would make the data in these databases more similar to those in good observational studies would eliminate the advantages of speed and cost (and therefore size) that make database research attractive.

Data-dredging—taking advantage of the convenience of large databases to test multiple hypotheses—is frequently raised as an issue in large database analysis (box 2-l). The issue is not unique to the analysis of administrative data (it can occur with RCTS as well, for instance), but it is of particular concern in this context.

The contribution of database analysis to comparisons of treatments maybe more appropriately assessed in terms of how it can contribute, rather than whether it is definitive (43a). Cross-design synthesis, a formal mechanism for incorporating the results of database research into a comparative assessment of treatments, was described in a General Accounting Office (GAO) report to Congress (188). This methodology attempts to formalize the use of database analysis as a complement to RCTS in comparing medical technologies.

[13] The use of newer statistical techniques is daunting to the average clinician. One concern that has been raised is that as statistical adjustment becomes more sophisticated, it will be poorly accepted by the providers whom researchers would like to influence. Researchers studying acute myocardial infarction note that the instrumental variable and hierarchical modeling techniques being developed there have been perceived as intuitive by a panel of clinicians who provide clinical support (127).

BOX 2-1: Testing Multiple Hypotheses

Large preexisting health databases are often convenient for testing multiple hypotheses about whether various characteristics of the patients or providers are related to particular outcomes. If enough combinations are examined, however, some of the characteristics will have closer-than-expected relationships to some of the outcomes by chance alone-i. e., they will be statistically significant. These relationships might be interpreted as important, even though the odds of finding at least one statistically significant relationship in every 14 such combinations are better than 50-50, even if there are no real relationships among the variables.

The practice of testing multiple hypotheses in search of one that is statistically significant is known as data-dredging. Whether this process has taken place may not be readily apparent to the reader, particularly inasmuch as the immensity of the databases invites the use of complex multivariable statistical techniques (138). Although multiple analyses of research datasets are common and often appropriate, the potential magnitude of the problem in large administrative databases grants it special importance (124).

Fortunately, the size of the databases makes statistical approaches—such as developing hypotheses in one half of a dataset and testing them in the other—feasible for addressing the problem (7,138). Moreover, because many similar databases are available, interesting results can be subsequently retested in independent databases. In addition, experienced database researchers develop analytical plans that focus on relationships suggested by previous research or theory, which decreases the likelihood of spurious results. Finally, as with the findings of any other form of research, the results of database analyses should be examined in the context of a much larger body of research.

SOURCE: Jeff Whittle, 1995.

ADJUNCTS TO OTHER RESEARCH METHODS

Applications

Research that combines primary data collection with the analysis of claims and discharge abstract databases reaps the advantages of both methods. The use of large, population-based databases as sampling frames, for example, facilitates the identification of representative samples. One of the original purposes of the National Cancer Institute's SEER program was to provide researchers with a tool: the case-finding capability of the SEER network (210) of cancer registries, each of which lists all cancers diagnosed in residents of a particular area. Data from the Professional Activity Study conducted by the Commission on Professional and Hospital Activities (CPHA) have been analyzed to identify potential cases for case-control studies of unusual occurrences, such as myocardial infarction following the use of oral contraceptives (96,97). Because myocardial infarction in women of childbearing age is rare, a single center could not have accumulated an adequate number of cases for a study. CPHA, however, collects data from thousands of North American hospitals and has a database of more than 150 million discharge abstracts (96).

Medicare eligibility files have been used to define a representative cohort of elderly individuals. The Medicare hospitalization file has been used to identify representative samples of discharge ab-

stracts for studies of coding (79,80), for assessments of the quality of care before and after the introduction of Medicare's prospective payment system for hospital inpatient care (100), and for studies of the outcomes experienced by patients who had suffered myocardial infarction, stroke, pneumonia, or congestive heart failure (30). Before studying the appropriateness of various procedures, researchers have used Medicare files for professional claims to identify cohorts of patients who have undergone various procedures (22). Claims databases can also be used as sampling frames for pseudorandomized trials that take advantage of the varied treatment assignments created by regional differences in the treatment of common conditions (120).

Issues

One concern about using databases as sampling frames is that researchers may have difficulty obtaining data about the patients whose cases have been identified. The researchers who studied the effects of the Medicare prospective payment system on the care provided to the program's beneficiaries, however, managed to obtain the medical charts for 96.2 percent of the patients in a sample identified from a Medicare claims database (100). The analysts who conducted the coding studies of 1985 and 1988 obtained 99.6 percent and 91.8 percent, respectively, of the charts of the patients they had identified (79,80).

Other researchers have used Medicare data to identify representative samples of hospitalized patients and then contacted them or their health care providers to obtain additional information. Researchers studying cataract surgery are using the Medicare database to find a sample of cases for further study of how posterior capsulotomy following the surgery is associated with retinal detachment (174). The researchers will contact providers to obtain information regarding factors (e.g., the length of the eyeball) that place patients at risk for retinal detachment but that cannot be determined from the claims data. These data will permit the researchers to control for differences in those variables, providing stronger observational evidence that the posterior capsulotomy itself increases the risk of retinal detachment. PORT researchers studying prostate disease and total knee replacement have contacted patients to ask about their levels of functioning after certain procedures have been performed on them (52,145).

Another major concern is that the patients' inclusion in the databases is involuntary. Consequently, a request to participate in a study can be an unexpected imposition. Nonetheless, researchers who have taken this approach have found that the individuals are generally willing to participate. These researchers believe that the privacy mechanisms currently in place are adequate to protect patients' confidentiality and their freedom to choose whether to participate in a study. (The Institute of Medicine has recently issued a report with recommendations regarding national policy on the conflict between patients' privacy and databases' usefulness.)

If patients are willing participants, many health care providers are not. Some providers will not comply with requests for records or participate in studies of their decisionmaking processes (101). The fact that providers who decline to participate may be systematically different corrupts the very generalizability that makes the use of databases as sampling frames so attractive. Researchers at the RAND Corporation have described methods to increase participation (108), but the methods are costly and do not result in participation by 100 percent of the providers.

Large databases can also be used in conjunction with primary data collection to provide followup

[14] Each patient receives a letter from the Health Care Financing Administration several weeks before the researchers plan to contact them. The letter tells patients about the study and allows them to indicate that they prefer not to be contacted. A phone number is provided for patients who have questions. Of 1,750 patients who received letters as part of one PORT project, two called and expressed grave concern that this step had been taken without their consent (145).

for populations that are well characterized by the primary data collection. Researchers at the National Center for Health Statistics, for example, have used personal identifiers to link individuals who are included in several surveys (e.g., the National Health Interview Survey Supplement on Aging) (94,187) with information about the individuals in Medicare databases. Using personal identifiers that allow linkage between RCTS and the National Death Index is an inexpensive way to improve long-term followup of the participants' vital status (29). Linkage to the Medicare claims database could also provide information regarding the need for hospitalization for specific diagnoses, should yield estimates of the costs of subsequent care, and might improve researchers' chances of contacting patients directly. Unfortunately for researchers, people under the age of 65 have no identifiers other than their Social Security numbers and have no population-wide insurance system that could (like Medicare) be used to track medical events over time.

Variant approaches that use claims databases both for identifying samples and for conducting followup are surveys of the Medicare population. The Medicare Current Beneficiary Survey gathers personal data in interviews with samples of beneficiaries, then obtains followup information from the Medicare claims data. More than 12,000 beneficiaries were surveyed in 1991, the first year in which data were collected (149,1 82). The Medicare Beneficiary Health Status Registry uses a mailed survey to obtain data on a spectrum of issues affecting health, including lifestyle risk factors, functional status, medical history prior to Medicare eligibility, sociodemographics, and quality of life (1 26). Neither the Medicare Current Beneficiary Survey nor the Medicare Beneficiary Health Status Registry has existed long enough for an assessment of its utility.

OTHER ISSUES

The issues in database research vary with the technology under consideration, the database, and the focus of the assessment. Thus, coding inaccuracies are much more important for a study of myocardial infarction (where the error rates are high, perhaps in a biased pattern) than for a study of lung cancer (where the error rates are low). Similarly, the inability to distinguish the reason for a procedure in outpatient Medicare data might affect a study of mammography (a procedure that could just as easily be done for diagnosis or screening as for a workup for known cancer), but would not affect a study of cataract extraction (a procedure that is generally performed only for cataracts related to existing or anticipated visual impairment).

Enhancing Databases

As the use of databases for assessing health care has grown, so has the realization that the existing databases are often inadequate for the proposed uses. This has stimulated interest in designing databases that are better suited for the analyses. Strategies for doing so include not only improving and augmenting the data but, in some cases, even designing entirely new databases (box 2-2).

Collecting More Data

A number of studies have shown that clinical data beyond the ICD-9-CM-coded discharge abstract data can explain the differences in the resources used for patients in identical diagnosis-related groups (the clinical grouping categories used as a basis for Medicare payment) (18,36,75,139). By collecting this data in multiple hospitals at multiple locations, proponents of these systems have shown that it is feasible for the data to be collected on a large scale (3,88). The addition of just three clinical variables not reflected in the discharge abstract data markedly improves the ability of a model based on ICD-9-CM data to predict mortality following CABG surgery (61).

The costs of such data collection, however, can be high. The experience of the SEER program is instructive. Early data collected in SEER registries included detailed data regarding the extent to which each patient's cancer had spread. Quality-control activities disclosed that the reliability of the data regarding fine gradations in stage was limited, but the data were very accurate for distin-

BOX 2-2: Creating a New Kind of Database: Community Health Management Information Systems

An ambitious method of addressing the problem of limited data is to create an entirely new database that both links and augments existing data sources. The Hartford Foundation's Community Health Management Information System project is an example of an effort that has begun the process of developing a communitywide database to enhance both the quality of care and the ability to assess the effects of medical treatment.

The foundation has organized community leaders to consider the advantages of improved health data systems in several cities around the country, with varying degrees of support from local governments. Which elements the database should include, how the data should be gathered, and how the quality of the data should be maintained are all decisions that will be made explicitly as the database is designed.

SOURCE: Jeff Whittle, 1995.

guishing coarser gradations. The coding system was designed to provide optimally useful scientific information on stage, but practical difficulties in obtaining accurate coding limited the use of the data, which were gathered at considerable cost (91). The difficulty of precise coding in this research-oriented, single-disease database should give pause to those who want to initiate more detailed collection in other large databases.

Pennsylvania now requires that all discharge abstracts include the key elements necessary for determining the MedisGroups severity levels (12,86). Researchers in the pneumonia PORT have used the nationwide MedisGroups database to analyze predictors of length-of-stay and mortality among patients admitted with pneumonia (42,43); they will use the statewide Pennsylvania MedisGroups database in the future.

HCFA has had an active interest in a similar set of clinical data, the Uniform Clinical Data Set (UCDS), for several years (109). The UCDS began as a set of more than 1,700 clinical data elements that could be collected by reviewing the charts associated with a subset of Medicare discharge abstracts. As time has gone by, the UCDS concept has evolved to a more flexible model that entails collecting different sets of data for different clinical entities (37).

A careful assessment of the experience of those who use the databases with supplemental clinical data may provide future guidance regarding the overall usefulness of a number of variants of this technique.

Including a Health Status Measure

Health status is an important outcome that is unavailable in current databases. Because most treatments are intended to improve the patients' health rather than to prolong their lives, death as an outcome of treatment is likely to be inapplicable in many situations.

Some observers have suggested incorporating a health status measure in the claims records maintained by Medicare. Selecting a measure appropriate for all patients, however, is problematic. The best measure for assessing the benefits of total hip replacement is certainly not the best measure for assessing the benefits of cataract surgery. Moreover, the important outcomes of surgery often cannot be determined at the time of hospital discharge. The reduction in pain following a hip replacement is best assessed after the patient has been discharged, and patients who undergo radical prostatectomy cannot be expected to assess their sexual functioning before leaving the hospital. One approach would be to allow researchers to

contact the patients later to obtain the specific functional data of interest. This has been done by research groups studying the effects of prostatectomy and total knee replacement.

The experience of the Medicare Beneficiary Health Status Registry may provide guidance regarding whether and how health-status outcomes might be added to claims and discharge abstract databases.

Linking Databases

Where the data in a single database are limited, researchers can sometimes combine two databases. In a study comparing open prostatectomy with TURP in the treatment of BPH, the only clinical detail available from the claims database was coded with ICD-9-CM, but some of these patients could be linked to a list of patients for whom an anesthesiologist had carried out preoperative risk assessments (161).

This approach can sometimes provide useful additional data. Epidemiologists link databases when they determine vital status by comparing the names of study participants with information from death registries, driver's-license agencies, and telephone books. By carefully reviewing each linked piece of data, the researchers can be reasonably certain that the data refer to the same individuals (9).

Database linkages can provide information beyond vital status. Researchers have linked state discharge data and cancer registry data from New Jersey to study differences in the results obtained by breast cancer patients with varying insurance coverage (4). A linked Medicare-SEER database, currently under development, will provide information about patients' treatment and the costs of their care from Medicare and detailed information about their cancer from SEER. Linkages between Medicare and Medicaid allow researchers to obtain information about the use of prescription drugs from Medicaid and longitudinal followup from Medicare (147,148).

The experience of several investigators demonstrates the feasibility and potential usefulness of this approach. The SEER-Medicare link—using each patient's name, sex, Social Security number, date of birth, and date of death—has identified the Medicare claims record for nearly 95 percent of the individuals over the age of 64 who are in the SEER cancer registry. In another case, Medicare records were identified for 85 percent of the men over 64 who used VA facilities in the Northeast over a four-year period. The researchers were able to study the degree to which this cohort used medical services provided by entities other than the VA—an accomplishment that has facilitated more accurate interpretations of studies of the VA administrative database (49).

Database linkages are not without problems, however. Most practically, linked studies require access to two (or more) databases, which doubles (or more) the cost of acquisition, the potential for violations of privacy, and the amount of data cleaning that is needed. In addition, different databases may use different definitions for similar concepts. For example, the coding for cancer surgeries in the SEER database differs from the ICD-9-CM system in the Medicare hospital database and from the Current Procedural Terminology (CPT) ***system in*** the professional claims database.[15]

Other problems of linkage are more technical. To a greater or lesser extent, all linkages are probabilistic—that is, the researcher identifies pairs of members of each database that have a certain probability of being the same persons (two records that share the same sex, the same last name, the same date of birth, and the same maternal last name represent the same person with a certain likelihood). The use of unique identifiers, such as social security numbers, can allow very high confidence about a match (160), but they often fail to be truly unique. For example, women who are eligible for Medicare because of their husbands' eligibility use their husbands' social security num-

[15] The Agency for Health Care Policy and Research is working to address these issues (25).

bers (although each beneficiary's own number will be included in the Medicare claims database in the future). In addition, many older individuals have been assigned more than one social security number (104). [16]

Methods of increasing linkage, and the confidence with which probabilistic matches can be regarded as true matches, are the subjects of active research. Issues include increasing linkage rates, enhancing the accuracy of links, providing estimates of the accuracy of links, dealing with uncertain linkages, and minimizing the computational burden of matching records between large databases {78,159,1 81,1 85).

I The Electronic Medical Record

As expanded databases become more complete, researchers will have access to more of the data that are in patients' charts. Taken to its extreme, this concept would result in the complete computerization of medical records, which could provide researchers with access to all the information that is generated during the patients' hospital stays.

Researchers at Beth Israel Hospital in Boston have used a computerized hospital database for more than 10 years and have demonstrated its ability to easily supplement the information found in the coded discharge abstracts. This does not eliminate the concern that risk adjustments made with these data may be biased because the data are obtained for nonrandom reasons (166). The lack of uniformly applied diagnostic criteria remains a potential source of error in any database that does not impose definitions for the diagnoses of interest, whether the medical record is coded or electronic. Moreover, certain data that a researcher might be anxious to have (e.g., scores on a particular functional status scale) are unlikely to appear anywhere in the medical record. Of course, many of these concerns apply whether the data about patients are obtained from a readily used, computerized practice database or by a laborious review of charts (the traditional method).

The electronic medical record has many potential uses for research into the effectiveness of health care. Inasmuch as more data are available in an electronic format, however, researchers should find it easier to retrieve information about samples of patients who have been identified through claims databases. The electronic medical record, therefore, will probably complement, rather than replace, large databases as tools for evaluating medical technologies.

Retrieving Primary Data

One alternative to adding information to the database is to allow researchers to retrieve the required data directly. Access to primary data allows theresearcher more flexibility in choosing data elements, as well as direct control over the quality of the data collection. With the ability to contact patients or providers, researchers could even identify data that are not recorded in the charts. This approach raises difficult questions of privacy (inasmuch as the patients may not be asked for permission to use their charts), logistics (because some databases reflect admissions throughout the country), and selection bias (because some charts-which are likely to be the unusual ones—will not be found) (168).

Several studies have demonstrated high rates of record retrieval when Medicare databases were used. Moreover, both the prostate and total-knee-replacement PORT teams have found that when individuals who are identified through database analyses are contacted directly, they cooperate with research efforts. In addition, HCFA and individual researchers have developed methods that facilitate research with appropriate concern for privacy rights. There is little experience with similar research using private or state databases, perhaps because concerns about the potential for law-

[16] The vast majority of these people use only one of the numbers; they acquired the others early in the history of Social Security, when some people thought a different number was needed for each job.

suits provide strong incentives for private insurers to keep researchers from contacting the individuals in the insurers' databases.

Other Possibilities for Improvement

Unique Identifiers

Probably the most common concern of database researchers is their inability to get a complete picture of medical services used over time, except from the Medicare claims data. The root of the problem is the lack of unique identifiers for patients in most other databases. The identifiers in insurance databases are unique, but patients move into and out of the systems relatively frequently, which hinders followup. Moreover, insurance databases generally do not include representative samples of the population.

Unique identifiers would provide several benefits:

- Researchers could more easily follow patients through time, between providers, and in various settings.
- All health care services covered by a database could be linked.
- Linkage would be simplified.
- A representative sample could be generated by starting with the list of all active unique identifiers.

Improving Data Accuracy

Improving the accuracy, completeness, and reliability of medical data is widely believed to be a good thing, for obvious reasons.

There are several potential mechanisms to facilitate improvements. For example, the computers used in the data entry systems could be programmed to reject out-of-range values, inconsistent codes, and nonspecific codes as the data are entered. Moreover, the programs could be altered to prevent the submission of incomplete data, to incorporate prompts regarding common errors, and to minimize transcription errors. Precise guidelines on appropriate coding could also help by making coding more consistent. Although the coding manuals used by all coders contain identical codes, the ambiguities in the ICD-9-CM system lead to wide variations in how coders in different sites apply the codes. Another useful mechanism would be to provide financial incentives for accurate and complete coding. Unfortunately, it is more difficult to envision a reward system than a system that imposes payment delays or other financial penalties for coding that fails to meet standards.

Regardless of any improvements, the data will always contain errors. As with research involving primary data collection, assessing the quality of data is important to understanding the results of database research. Ideally, researchers should be able to assess the quality of specific data elements in the specific database they use. With the exception of reports on Medicare hospital-claims data, however, few studies regarding the quality of data have been published.

Improving the Coding Systems

Inasmuch as the ICD-9-CM coding system itself has been blamed for much of the difficulty with using claims data, changing the coding system might enhance the databases' usefulness. Obvious gaps that could be filled include information about whether diseases were present or procedures were performed on the patient's right side, left side, or both; whether conditions existed at the times of admission or developed during the hospitalization; and how severe conditions are. Like adding new fields of data to be collected, however, increasing the complexity of the coding system adds to the data-collection burden. In New York, a binary code to designate whether a condition was present at admission has been introduced but has not been reliably implemented (61).

The ICD-9-CM is under continuous revision. In addition to HCFA, the entities most concerned are special-interest groups that wish to make certain conditions more precisely identifiable. Reliable, regular communication with persons knowledgeable about the data needs of medical researchers could provide guidance about how proposed coding changes might affect the usefulness of coded data for their research. For example, AHCPR staff made suggestions regarding

changes that could make the coding of incontinence and prostatic diseases more useful for evaluating the effects of treatment (131).

National standardization of HCFA'S coding system for professional claims might facilitate the use of professional service claims in the Medicare database.

Exchanging Information About Database Research Methods

Information about how the data in a database are collected, verified, and stored is crucial for understanding and using a database. Although the accuracy and completeness of specific data elements in particular databases may have been studied, the results often are not published and summary documents are unavailable, which means that other researchers must spend time and money rediscovering or recreating vital pieces of information. Increasing the availability of such information is a major focus of AHCPR'S Office of Science and Data Development (184), but researchers are reluctant to write summaries of their findings about the databases, because the subject is too arcane for publication and the writing is too time-consuming to undertake without strong incentives.

Although there are many published and unpublished studies of the quality of the information in various databases, the need for current information about the accuracy of data never ceases. Continuous changes in coding incentives and conventions can affect coding accuracy in ways that are unpredictable in magnitude, if not direction. Because the accuracy of databases varies considerably, depending on which conditions are being studied, information on the specific codes of interest must be used. This usually means that the researchers must carry out coding studies that specifically target the issue at hand, because even the large (more than 7,000 charts) validation studies carried out in connection with the change in Medicare payment included fewer (usually much fewer) than 100 patients with any one condition. Ideally, the studies should be published, but they frequently are not.

Methodologic Advances

Certain aspects of methodology are particularly important for research using claims and discharge abstract data. For example, better case-mix adjusters based only on ICD-9-CM data would be extremely useful. This area has been active, with several proprietary and open systems available. At present, there is no consensus as to the most useful method, although different methods will probably prove best for different applications.

Newer statistical methodologies, including hierarchical modeling (129), are being explored for use in adjusting. Of course, improvements in statistical techniques do not eliminate the requirement for accurate, reliable data on important potential confounders.

Analytical Costs

Compared with research that entails primary data collection, database analysis is inexpensive (92,175). Researchers incur fewer expenses for collecting claims data, for example, which has already been gathered for billing purposes. Moreover, because the patients are receiving routine clinical care and the clinicians are paid as part of routine practice, the study does not need to fund the patients' care (202). Nonetheless, a number of the costs involved in preparing the data for use in research should be considered.

First, the data must be acquired from whomever collects them. The charges are generally low compared with those for primary data collection, but they are not insignificant. The costs of the public use tapes that include 100 percent samples of Medicare hospital discharge data with linkable identifiers, for example, are $6,120 for each year of data (182). Pennsylvania provides state discharge abstracts, supplemented with the admission and followup MedisGroups severity scores,[17] for one cent per discharge plus the computer costs of se-

[17] MedisGroups is the trade name for a system that uses clinical data from hospital records to generate scores that predict individual patients' costs and outcomes. These data are also used to detect clinical deterioration that occurred during the patients' hospitalization and that might signal problems with the quality of the care being delivered.

lecting the desired population. The cost of acquiring discharge abstract data enhanced with the more than 100 clinical data elements that go into the MedisGroups scoring system was 15 cents per record for the pneumonia PORT.

Elements of particular interest to researchers are sometimes poorly documented, and the researchers incur additional costs in time and computer resources while discovering the weaknesses in the data and validating the data elements. Researchers who wish to link two or more databases must pay the costs of acquiring and examining each database separately and then performing the linkage. Furthermore, the statistical analyses are generally more expensive to perform in studies using large databases than in studies using primary data collection from relatively small numbers of individuals.

As efforts are made to enhance the databases, the costs of data collection itself may become a significant factor. Gathering data solely for a database is not inexpensive. The Connecticut Tumor Registry, for example, lists nearly 15,000 new cancer cases each year (and follows them up) at a cost of approximately $1 million (170). If elements are added to routine billing data for research purposes, the added costs should be assigned to the research. [18] As part of an effort to provide risk-adjusted data on costs and outcomes in state hospitals, the Pennsylvania Health Care Cost Containment Commission (HCCCC) has, since 1986, required most of the state's hospitals to gather the clinical data (Key Clinical Findings) needed for determining the MedisGroups severity scores when patients are admitted and again during their hospital stays (12,86).

Pennsylvania's experience illustrates the potential costs of such data collection. The HCCCC has a budget of $2 million, somewhat over half of which is allocated to data analysis (38). The added cost to the hospitals is significant: researchers calculated the cost of collecting these dataat$13.90 per discharge for the first year (88), which—in a state with approximately 2 million discharges annually—meant that the total cost of compliance approached $25 million. Anecdotal evidence suggests that the costs decrease by $1 or $2 per discharge as the process becomes routine, but a statewide cost in excess of $20 million annually appears likely. MediQual, which markets the MedisGroups system, estimates the time commitment at 20 minutes per medical chart and 30 minutes per surgical discharge abstract (136).

The potential usefulness of the MedisGroups data for quality assurance had led 15 percent of the hospitals in Pennsylvania to collect this data prior to the HCCCC'S mandate. Moreover, the value of the data for nonresearch activities should be considered when evaluating the cost. Nonetheless, if data requirements are imposed on hospitals to support research activities, a careful consideration of the financial implications for all parties concerned is warranted.

Researchers working with the PORTS report varying commitments of time and resources to database analyses, but all agree that the amount of time is both greater than expected and substantial. An estimate of the actual cost is difficult to obtain, because time allotted by salaried investigators is a major variable. Depending on the PORT, between 10 percent and half of the resources committed to the overall project have been devoted to database analyses (102).

Several things are likely to reduce the costs of database analysis in the future. Information about which elements in databases are reliable will become available from the researchers who are currently exploring them, which will save future re-

[18] Because health maintenance organizations (HMOs) are not now required to generate claims as part of the billing process, any costs that HMOs incur because of requirements that they generate similar data might be considered research costs as well. Anecdotally, many HMOs gather fairly extensive utilization data as part of their internal quality control activities, so this concern may be only theoretical.

searchers a lot of time.[19] Databases may become available in already linked form, allowing the cost of linkage to be paid just once. After two databases have been linked, the cost of updating them will probably decrease. AHCPR has been active in trying to identify and remove the obstacles to the efficient use of databases (185).

REFERENCES

1. Abrams, H. B., Detsky, A. S., Roos, Jr., L.L., et al., "Is There a Role for Surgery in the Acute Management of Infective Endocarditis? A Decision Analysis and Medical Claims Database Approach," *Medical Decision Making 8: 165-174,* 1988.
2. American Medical Association, *CPT Physicians' Current Procedural Terminology* (Chicago, IL: American Medical Association, 1991).
3. Averill, R. F., McGuire, T. E., Manning, B. E., et al, "A Study of the Relationship Between Severity of Illness and Hospital Cost in New Jersey Hospitals," *Health Services Research* 27:587-606; *1992.*
4. Ayanian, J. Z., Kohler, B. A., Abe, T., et al., "The Relation Between Health Insurance Coverage and Clinical Outcomes Among Women with Breast Cancer," New *England Journal of Medicine* 329:326-331, *1993.*
5. Beebe, G. W., "Long-Term Follow-Up Is a Problem," *American Journal of Public Health* 73:245-246, 1983.
6. Beta-Blocker Heart Attack Trial Research Group, "A Randomized Trial of Propranolol in Patients with Acute Myocardial Infarction I.: Mortality Results," *Journal of/he American Medical Association* 247: *1707-1714, 1982.*
7. Blumberg, M. S., "Potentials and Limitations of Database Research Illustrated by the QMMP AMI Medicare Mortality Study," *Statistics in Medicine* 10:637-646, *1991.*
8. Boissel, J. P., "Impact of Randomized Clinical Trials on Medical Practices," *Controlled Clinical Trials* 10 (suppl.):120S-134S, 1989.
9. Boyle, C. A., and Decoufle, P., "National Sources of Vital Status Information: Extent of Coverage and Possible Selectivity of Reporting, ''American Journal of Epidemiology* 131:160-168, *1990.*
10. Braunwald, E., *Heart Disease: A Textbook of Cardiovascular Medicine* (Philadelphia, PA: W.B. Saunders, 1992).
11. Brem, H., John Hopkins Hospital, Baltimore, MD, personal communication, January 1990.
12. Brewster, A. C., Karlin, G. B., Hyde, L. A., et al., "MEDISGROUPS: A Clinically Based Approach to Classifying Hospital Patients at Admission," *Inquiry* 22:377-387, *1985.*
13. Brook, R. H., Kosecoff, J. B., Park, R. E., et al., "Diagnosis and Treatment of Coronary Disease: Comparisons of Doctors' Attitudes in the USA and the UK," *Lancet* 1:750-753, 1988.
14. Bunker, J. P., "Surgical Manpower: A Comparison of Operations and Surgeons in the United States and in England and Wales," New *England Journal of Medicine 282:* 135-*144, 1970.*
15. Byar, D. P., "Why Data Bases Should Not Replace Randomized Clinical Trials," *Biometrics* 36:337-342, *1980.*
16. Byar, D. P., "Problems with Using Observational Databases to Compare Treatments," *Statistics in Medicine* 10:663-666, *1991.*

[19] At this time, there is no reliable source for such information. Because it is of interest to only the few people carrying out database research, it is unlikely to be published in widely circulated medical or health services research journals. AHCPR has facilitated the exchange of such information by sponsoring conferences that provide opportunities for database researchers to exchange information (217). The products of these conferences serve as valuable resources. AHCPR has also established the Use of Claims Data Work Group, which includes many of the PORT investigators who analyze claims.

17. Byar, D.P., Schoenfeld, D.A., Green, S. B., et al., "Design Considerations for AIDS Trials," New *England Journal of Medicine* 323:1343-1348, *1990.*
18. Calore, K.A., and Iezzoni, L. I., "Disease Staging and PMCs: Can they Improve DRGs?" *Medical Care* 25:724-737, *1987.*
19. Devita, Jr., V.T., Hellman, S. and Rosenberg, S.A. (eds.), *Cancer: Principles and Practice of Oncology* (Philadelphia, PA: J.B. Lippincot Company, 1993).
20. CASS Principal Investigators and Their Associates, "Coronary Artery Surgery Study (CASS): A Randomized Trial of Coronary Artery Bypass Surgery. Comparability of Entry Characteristics and Survival in Randomized Patients and Nonrandomized Patients Meeting Randomization Criteria," *Journal of the American College of Cardiology 3:* 114-128, 1984.
21. Chassin, M.R., Brook, R. H., Park, R. E., et al, "Variations in the Use of Medical and Surgical Services by the Medicare Population," New *England Journal of Medicine* 314 (30):285-290, *1986.*
22. Chassin, M. R., Kosecoff, J., Park, R.E., et al., "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? A Study of Three Procedures," *Journal of the American Medical Association 258* (18):2533-2537, *1987.*
23. Clearly, P. D., Greenfield, S., Mulley, A.G. et al., "Variations in Length of Stay and Outcomes for Six Medical and Surgical Conditions in Massachusetts and California," *Journal of the American Medical Association* 26:73-79, 1991.
24. Cobb, L. A., Thomas, G. I., Dillard, D. H., et al., "An Evaluation of Internal-Mammary - Artery Ligation by a Double-Blind Technique," New *England Journal of Medicine* 260:1115-1118, *1959.*
25. Cohen, E., Bemier, D., Tam, S., et al., "Data Quality and DRGs: An Assessment of the Reliability of Federal Beneficiary Discharge Data in Selected Manhattan Hospitals," *Journal of Community Health* 10:238-246, 1985.
26. Coleman, M., "Chemotherapy for Large-Cell Lymphoma: Optimism and Caution," *Annals of Internal Medicine* 103:140-142, *1985.*
27. Cohen, M. F., "Clinical Research Databases—A Historical Review," *Journal of Medical Systems* 14:323-344, *1990.*
28. Concato, J., Horwitz, R. I., Feinstein, A. R., et al., "Problems of Comorbidity in Mortality After Prostatectomy," *Journal of the American Medical Association 267:1077-1082, 1992.*
29. Curb, J. D., Ford, C.E., Pressel, S., et al., "Ascertainment of Vital Status Through the National Death Index and the Social Security Administration," *American Journal of Epidemiology* 121:754-766, *1985.*
30. Daley, J., Jencks, S., Draper, D., et al., "Predicting Hospital-Associated Mortality for Medicare Patients: A Method for Patients with Stroke, Pneumonia, Acute Myocardial Infarction, and Congestive Heart Failure," *Journal of the American Medical Association* 260:3617-3624, 1988.
31. Demlo, L. K., and Campbell, P. M., "Improving Hospital Discharge Data: Lessons from the National Hospital Discharge Survey," *Medical Care 19: 1030-1040, 1981.*
32. Doughty, A., Nash, S.1., and Gift, D. A., "Deployment and Utilization of MR Imaging in Michigan: Observations of a Statewide Data Base," *Radiology* 185:53-61, *1992.*
33. Dyck, F.J., Murphy, F.A., Murphy, J. K., et al, "Effect of Surveillance on the Number of Hysterectomies in the Province of Saskatchewan," New *England Journal of Medicine* 296:1326-1328, *1977.*
34. Eddy, D. M., and Billings, J., "The Quality of Medical Evidence: Implications for Quality Care," *Health Affairs* 7(1):19-32, spring 1988.
35. Eggers, P.W., Connerton, R., and McMullan, M., "The Medicare Experience with End-Stage Renal Disease: Trends in Incidence,

Prevalence, and Survival," *Health Care Financing Review* 5:69-88,1984.

36. Eisenberg, B.S., "Diagnosis-Related Groups, Severity of Illness and Equitable Reimbursement Under Medicare," *Journal of the American Medical Association* 251(5): *645-646, 1984.*
37. Ellerbeck, E., U.S. Department of Health and Human Services, Health Care Financing Administration, Office of Quality Assurance, Baltimore, MD, personal communication, June 1993.
38. Essa, E.J., Health Care Cost Containment Commission, Harrisburg, PA, personal communication, June 1993.
39. Farow, D. C., Hunt, W. C., and Samet, J. M., "Geographic Variations in the Treatment of Localized Breast Cancer," New *England Journal of Medicine* 326(17):1097-1101, *1992.*
40. Feinstein, A.R., "ICD, POR, and DRG: Unsolved Scientific Problems in the Nosology of Clinical Medicine," *Archives of Internal Medicine* 148:2269-2274, *1988.*
41. Feinstein, A.R., Sosin, D. M., and Well, C. K., "The Will Rogers Phenomenon: Stage Migration and New Diagnostic Techniques as a Source of Misleading Statistics for Survival in Cancer," New *England Journal of Medicine* 312:1604-1608, *1985.*
42. Fine, M.J., Hanusa, B. H., Lave, J.R., et al., "Variation in Length of Stay in Patients Hospitalized with Community-Acquired Pneumonia" (abstract), *Clinical Research 39: 142A, 1991.*
43. Fine, M.J., Hanusa, B. H., Singer, D.E., et al., "Validation of a Pneumonia Mortality Risk Index Using the MedisGroups Comparative Hospital Database" (abstract), Clinical Research 39: 142A, 1991.

43a. Fisher, E. S., Dartmouth Medical School, Hanover, NH, personal communication, 1993.

#. Fisher, E. S., Baron, J. A., Malenka, D.J., et al., "Overcoming Potential Pitfalls in the Use of Medicare Data for Epidemiologic Research," *American Journal of Public Health* 80:1487-1490, *1990.*

45. Fisher, E. S., Baron, J.A., Malenka, D.J., et al., "Hip Fracture Incidence and Mortality in New England," *Epidemiology* 2:116-122, *1991.*
46. Fisher, E. S., Malenka, D.J., Wennberg, J.E., et al., "Technology Assessment Using Insurance Claims. Example of Prostatectomy," *International Journal of Technology Assessment in Health Care 6: 194-202, 1990.*
47. Fisher, E. S., Whaley, F. S., Krushat, W. M., et al., "The Accuracy of Medicare's Hospital Claims Data: Progress Has Been Made but Problems Remain," *American Journal of Public Health* 82:243-248, *1992.*
48. Fisher, R. I., Gaynor, E.R., Dahlberg, S., et al., "Comparison of a Standard Regimen (CHOP) with Three Intensive Chemotherapy Regimens for Advanced Non-Hodgkin's Lymphoma," New *EnglandJournal of Medicine 328:1002-1006, 1993.*
49. Fleming, C., Fisher, E. S., Chang, C.H., Bubolz, T. A., Malenka, D.J., "Studying Outcomes and Hospital Utilization in the Elderly: The Advantages of a Merged Database for Medicare and Veterans Affairs Hospitals," *Medical Care* 30:377-391, *1992.*
50. Fode, N. C., Sundt, T. M., Robertson, J.T., et al., "Multicenter Retrospective Review of Results and Complications of Carotid Endarterectomy in 1981," *Stroke* 17:370-376, *1986.*
51. Ford, E., Cooper, R., Castaner, A., et al., "Coronary Arteriography and Coronary Bypass Survey Among Whites and Other Racial Groups Relative to Hospital-Based Incidence Rates for Coronary Artery Disease: Findings From NHDS," *American Journal of Public Health* 79:437-440, *1989.*
52. Fowler, F. J., University of Massachusetts, Boston, MA, personal communication, October 1992.
53. Fowler, F.J., Wennberg, J.E., Timothy, R. P., et al, "Symptom Status and Quality of Life

Following Prostatectomy," *Journal of the American Medical Association 259:3018-3022, 1988.*

54. Friedman, L. M., Furberg, C.D., and DeMets, D. L., *Fundamentals of Clinical Trials* (Littleton, MA: PSG Publishing Company, Inc., 1985).
55. Fries, J.F., "Advances in the Management of Rheumatic Disease: 1965 to *1985,'' Archives of Internal Medicine 149:1002-1011,* 1989.
56. Gehan, E.A., "The Evaluation of Therapies: Historical Control Studies," *Statistics in Medicine* 3:315-324, *1984.*
57. Ginsberg, R.J., Hill, L.D., Eagan, R.T., et al., "ModemThirty-Day Operative Mortality for Surgical Resections in Lung Cancer," *Journal of Thoracic and Cardiovascular Surgery* 86:654-658, *1983.*
58. Glover, J.A., "The Incidence of Tonsillectomy in School Children," *Proceeding of the Royal Society of Medicine* 31:1219-1235, *1938.*
59. Glover, R.P., Davila, J. C., Kyle, R. H., et al., "Ligation of the Internal Mammary Arteries as a Means of Increasing Blood Supply To the Myocardium," *Journal of Thoracic and Cardiovascular Surgery* 34:661-678, *1957.*
60. Gordon, L. I., Barrington, D., Andersen, J., et al., "Comparison of a Second-Generation Combination Chemotherapeutic Regimen (m-BACOD) with a Standard Regimen (CHOP) for Advanced Non-Hopkin's Lymphoma," New *England Journal of Medicine* 327:1342-1349, *1992.*
61. Hannan, E.L., Kilburn, Jr., H., Lindsey, M. L., et al., "Clinical Versus Administrative Data Bases for CABG Surgery: Does It Matter?" *Medical Care* 30:892-907, *1992.*
62. Hannan, E.L., Kilburn, Jr., H., O'Donnell, J.F., et al., "Interracial Access to Selected Cardiac Procedures for Patients Hospitalized with Coronary Artery Disease in New York State," *Medical Care* 29:430-441, *1991.*
63. Harris, W. H., and Sledge, C. B., "Medical Progress: Total Hip and Total Knee Replacement (First of Two Parts)," New *England Journal of Medicine* 323:725-731,1990.
64. Hartz, A.J., Kuhn, E. M., Pryor, D. B., et al., "Mortality After Coronary Angioplasty and Coronary Artery Bypass Surgery (The National Medicare Experience)," *American Journal of Cardiology* 70: *179-185,* 1992.
65. Hatten, J., "Medicare's Common Denominator: The Covered Population," *Health Care Financing Review,* fall:53-64, 1980.
66. Held, P.J., Levin, N. W., Bovbjerg, R. R., et al., "Mortality and Duration of Hemodialysis Treatment," *Journal of the American Medical Association* 265(7):871-875, *1991.*
67. Hellman, S., and Hellman, D. S., "Of Mice But Not Men. Problems of the Randomized Clinical Trial," New *England Journal of Medicine* 324:1585-1589, *1991.*
68. Hine, L.K., Laird, N., Hewitt, P., et al., "Meta-Analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction," *Archives of Internal Medicine* 149:2694-2698, *1989.*
69. Hlatky, M.A., Stanford University, Palo Alto, CA, personal communication, July 1993.
70. Hlatky, M.A., "Using Databases to Evaluate Therapy," *Statistics in Medicine* 10:647-652, *1991.*
71. Hlatky, M. A., Califf, R. M., Harrell, F. E., et al., "Comparison of Predictions Based on Observational Data with the Results of Randomized Controlled Clinical Trials of Coronary Artery Bypass Surgery," *Journal of the American College of Cardiology* 11(2):237-*245, 1988.*
72. Holmes, M. D., Hodges, D., and Rich, J., "Racial Inequalities in the Use of Procedures for Ischemic Heart Disease," *Journal of the American Medical Association 261:3242-3243, 1989.*
73. Holtgrewe, H. L., "An American Urological Association Prospective, Randomized Clinical Trial in the Treatment of Benign Prostatic Hyperplasia," *Cancer* 70:351-354, *1992.*

74. Hori, R. Y., Lewis, J.L., Zimmerman, J.R., et al., "The Number of Total Joint Replacements in the United States," *Clinical Orthopaedics and Related Research* 132:46-52, *1978.*
75. Horn, S.D, Ashworth, M.A., "Adjusting DRGs for Severity of Illness," *Socioeconomics of Surgery*, I.M. Rutkow (cd) (St. Louis, MO: C.V. Moseby, 1989).
76. Horwitz, R. I., Viscoli, C. M., Clemens, J. D., et al., "Improved Observational Method for Studying Therapeutic Efficacy: Suggestive Evidence That Lidocaine Prophylaxis Prevents Death in Acute Myocardial Infarction," *Journal of the American Medical Association 246* (21):2455-2459, *1981.*
77. Horwitz, R. I., Viscoli, C. M., Clemens, J. D., et al., "Developing Improved Observational Methods for Evaluating Therapeutic Effectiveness," *American Journal of Medicine 89* (5):630-638, *1990.*
78. Howe, G. R., and Lindsay, J., "A Generalized Iterative Record Linkage Computer System for Use in Medical Follow-Up Studies," *Computers and Biomedical Research* 14:327-340, *1981.*
79. Hsia, D. C., Ahem, C. A., Ritchie, B.P., et al., "Medicare Reimbursement Accuracy Under the Prospective Payment System, 1985 to 1988," *Journal of the American Medical Association* 268:896-899, *1992.*
80. Hsia, D. C., Krushat, W. M., Fagan, A. B., et al., "Accuracy of Diagnostic Coding for Medicare Patients Under the Prospective Payment System," *New England Journal of Medicine* 318:352-355, *1988.*
81. Hsia, D. C., Krushat, W. M., and Moscoe, L. M., "Epidemiology of Carotid Endarterectomies Among Medicare Beneficiaries," *Journal of Vascular Surgery* 16:201-208, *1992.*
82. Hughes, R. G., Hunt, S. S., Luft, H. S., "Effects of Surgeon Volume and Hospital Volume on Quality of Care in Hospitals," *Medical Care* 25:489-503, *1987.*
83. Hulley, S. B., Cummings, S.R., Browner, W. S., et al., *Designing Clinical Research: An Epidemiologic Approach* (Baltimore, MD: Williams and Wilkins, 1988).
84. Iezzoni, L. I., "Using Administrative Diagnostic Data To Assess the Quality of Hospital Care. Pitfalls and Potential of ICD-9-CM," *International Journal of Technology Assessment in Health Care* 6:272-281, *1990.*
85. Iezzoni, L. I., Foley, S. M., Daly, J., et al., "Comorbidities, Complications, and Coding Bias: Does the Number of Diagnosis Codes Matter in Predicting In-Hospital Mortality?" *Journal of the American Medical Association* 260:3159-3163, *1988.*
86. Iezzoni, L.I., Moskowitz, M. A., "A Clinical Assessment of MedisGroups," *Journal of the American Medical Association 260:3159-3163, 1988.*
87. Iglehart, J. K., "Canada's Health Care System" (part 1), New *EnglandJournal of Medicine* 315:202-208, *1986.*
88. Institute of Policy Research & Evaluation, "Hospitals' Response to Pennsylvania's Mandated Cost and Quality Data Reporting; Center for Health Policy Research: CHPR Research Report," The Pennsylvania State University, May 1990.
89. Institute of Medicine, *Reliability of Hospital Discharge Abstracts* (Washington, DC: National Academy of Sciences, 1977).
90. Institute of Medicine, *Reliability of Medicare Hospital Discharge Records,* (Washington, DC: National Academy of Sciences, 1977).
91. Jablon, S., Garfinkel, L., Gordis, L., et al., *Jablon Report,* 1981.
92. Javitt, J. C., "Outcomes of Eye Care from Medicare Data" (editorial), *Archives of Ophthalmology 109: 1079-1080, 1991.*
93. Javitt, J. C., Tielsch, J. M., Canner, J. K., et al., "National Outcomes of Cataract Extraction. Increased Risk of Retinal Complications Associated with Nd:YAG Laser Capsulotomy. The Cataract Patient Outcomes Re-

search Team," *Ophthalmology 99: 1487-97, 1992.*

94. Jekel, J.F., "Rainbow Reviews V: Recent Publications of the National Center for Health Statistics,' 'Journal *of Clinical Epidemiology 45:* 1185-1190, 1992.

95. Jencks, S.F., Williams, D. K., Kay, T.L., "Assessing Hospital-Associated Deaths from Discharge Data: The Role of Length of Stay and Comorbidities," *Journal of the American Medical Association 260:2240-2246, 1988.*

96. Jick, H., "The Commission on Professional and Hospital Activities-Professional Activity Study: A National Resource for the Study of Rare Illnesses," *American Journal of Epidemiology* 109:625-627, *1979.*

97. Jick, H., Dinan, B., and Rothman, K.J., "Oral Contraceptives and Nonfatal Myocardial Infarction," *Journal of the American Medical Association* 239:1403-1406, *1978.*

98. Kahn, C. N., "Policy Implications of Outcomes Research," *International Journal of Technology Assessment in Health Care* 6:295-296, *1990.*

99. Kahn, H.A., *An Introduction of Epidemiologic Methods (New* York, NY: Oxford University Press, 1983).

100. Kahn, K. L., Rogers, W. H., Rubenstein, L. V., et al., "Measuring Quality of Care with Explicit Process Criteria Before and After Implementation of the DRG-Based Prospective Payment System," *Journal of the American Medical Association* 264: *1969-1973,* 1990.

101. Kallich, J., RAND Corp., Santa Monica, CA, personal communication, May 1993.

102. Kapoor, W., University of Pittsburgh, Pittsburgh, PA personal communication, May 1993.

103. Keller, R. B., Soule, D. N., Wennberg, J. E., et al., "Dealing with Geographic Variations in the Use of Hospitals. The Experience of the Maine Medical Assessment Foundation Orthopaedic Study Group," *Journal of Bone and Joint Surgery* 72: *1286-1293, 1990.*

104. Kestenbaum, B., U.S. Department of Health and Human Services, Social Security Administration, Office of the Actuary, Baltimore, MD, persona! communication, July 1992.

105. Klimo, P., and Connor, J. M., "MACOP-B Chemotherapy for the Treatment of Diffuse Large-Cell Lymphoma," *Annals of Internal Medicine 149:* 1177-1181, 1989.

106. Knickman, J. R., and Foltz, A., "Regional Differences in Hospital Utilization: How Much Can Be Traced to Population Differences?" *Medical Care* 22:971-986, *1984.*

107. Kong, D. F., Lee, K. L., Harrell, Jr., F. E., et al., "Clinical Experience and Predicting Survival in Coronary Disease," *Archives of Internal Medicine* 102:596-602, *1985.*

108. Kosecoff, J., Chassin, M. R., Fink, A., et al., "Obtaining Clinical Data on Appropriateness of Medical Care in Community Practice," *Journal of the American Medical Associations* 258:2538-2542,1987.

109. Krakauer, H., "The Uniform Clinical Data Set," *Electiveness and Outcomes in Health Care,* K.A. Heithoff and K.N. Lohr (eds.) (Washington, DC: National Academy Press, 1990).

110. Krakauer, H., and Bailey, R. C., "Epidemiologic Oversight of the Medical Care Provided to Medicare Beneficiaries," *Statistics in Medicine* 10:521-540, *1991.*

111. Krakauer, H., Bailey, R. C., Skellan, K.J., et al., "Evaluation of the HCFA Model for the Analysis of Mortality Following Hospitalization," *Health Services Research 27:3:317-335, 1992.*

112. Kramer, M. S., and Shapiro, S. H., "Scientific Challenges in the Application of Randomized Trials," *Journal of the American Medical Association* 252:2739-2745,1984.

113. Kreuter, W., University of Washington, Seattle, WA, personal communication, July 1993.

114. Lave, J., Dobson, A., and Walton, C., "The Potential Use of Health Care Financing Ad-

ministration Data Sets for Health Care Services Research," *Health Care Financing Review* 5(1):93-98, *1983.*

115. Lawless, J. F., *Statistical Models and Methods for Lifetime Data (New* York, NY: John Wiley and Sons, 1982).
116. Lee, K. L., Pryor, D. B., Harrell, Jr., F.E., "Predicting Outcomes in Coronary Disease: Statistical Models versus Expert Clinicians," *American Journal of Medicine* 80:553-560, *1986.*
117. Lessler, J.T., Harris, B. S. H., *Medicaid Data as a Source for Postmarketing Surveillance Information: Final Report* (Chapel Hill, NC: Research Triangle Institute, 1984).
118. Lewis, C. E., "Variations in the Incidence of Surgery," New *England Journal of Medicine* 281:880-884, *1969.*
119. Localio, A. R., Lawthers, A.G., Bengtson, J. M., et al., "Relationship Between Malpractice Claims and Cesarean Delivery," *Journal of the American Medical Association* 269:366-373, *1993.*
120. Loft, A., Anderson, T.F., and Madsen, M., "A Quasi-Experiment Design Based on Regional Variations: Discussion of a Method for Evaluating Outcomes of Medical Practice," *Social Science and Medicine* 28: *147-154, 1989.*
121. Longo, D. L., deVita, Jr., V.T., and Young, R. C., "CHOP versus Intensive Regimens in Non-Hodgkin's Lymphoma," New *England Journal of Medicine* 329:580-581, *1993.*
122. Lu-Yao, G. L., McLerran, D., Wasson, J., et al., "An Assessment of Radical Prostatectomy: Time Trends, Geographic Variation and Outcomes," *Journal of the American Medical Association* 269:2633-2636, 1993.
123, Malenka, D.J., Roos, N., Fisher, E. S., et al, "Further Study of the Increased Mortality Following Transurethral Prostatectomy: A Chart-Based Analysis," *Journal of Urology* 144:224-227, *1990.*
124. Mantel, N., "Cautions on the Use of Medical Databases," *Statistics in Medicine* 2:355-362, *1983.*
125. McDonald, C.J., and Hui, S. L., "The Analysis of Humongous Databases: Problems and Promises," *Statistics in Medicine* 10:511-518, 1991.
126. McBean, M. A., U.S. Department of Health and Human Services, Health Care Financing Administration, Office of Research and Demonstrations, Baltimore, MD, personal communication, September 1993.
127. McNeil, B.J., Harvard University, Boston, MA, personal communication, August 1993.
128. McNeil, B.J., "Claims Data and Effectiveness: Acute Myocardial Infarction and Other Examples," *Electiveness and Outcomes in Health Care* K.A. Heithoff and K.N. Lohr (eds.) (Washington, DC: National Academy Press, 1990).
129. McNeil, B.J., Pederson, S. H., Gatsonis, C., "Current Issues in Profiling Quality of Care," *Inquiry* 29:298-307, *1992.*
130. McPherson, K., Wennberg, J. E., Hovind, O. B., et al., "Small-Area Variation in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway," New *England Journal of Medicine 307:1310-1314,* 1982.
131. Meads, S., U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD, personal communication, June 1993.
132. Mebust, W. K., Holtgrewe, H. L., Cockett, A.T., et al., "Transurethral Prostatectomy: Immediate and Postoperative Complications. A Cooperative Study of 13 Participating Institutions Evaluating 3,885 Patients," *Journal of Urology* 141:243-247, *1989.*
133. Medical Research Council, "Streptomycin Treatment of Tuberculous Meningitis," *Lancet 1:582-596,* 1948.
134. Melsey, K., Indiana University, Indianapolis, IN, personal communication, May 1993.
135. Miao, L. L., "Gastric Freezing: An Example of the Evaluation of Medical Therapy by Randomized Clinical Trials," *Costs, Risks and Benefits of Surgery,* J.P. Bunker, B.A.

Barnes, F. Mosteller (eds.) (New York, NY: Oxford University Press, 1977).
136. Monaghan, J., MediQual Systems, Westborough, MA, personal communications, July 1993.
137. Moses, L.E., "Framework for Considering the Role of Data Bases in Technology Assessment," *International Journal of Technology Assessment in Health Care* 6:183-193, *1990.*
138. Moses, L.E., "Innovative Methodologies for Research Using Databases," *Statistics in Medicine* 10:629-633, *1991.*
139. Munoz, E., Lausghlin, A., Regan, D. M., et al., "The Financial Effects of Emergency Department-Generated Admissions Under Prospective Payment Systems," *Journal of the American Medical Association 254:1763-1771, 1985.*
140. Nattinger, A. B., Gottlieb, M. S., Veum, J., et al., "Geographic Variation in the Use of Breast-Conserving Treatment for Breast Cancer," New *England Journal of Medicine* 326(17):1 *102-1107, 1992.*
141. Paradise, J. L., Bluestone, C. D., Bachman, R.Z., et al, "Efficacy of Tonsillectomy for Recurrent Throat Infection in Severely Affected Children: Results of Parallel Randomized and Non-Randomized Clinical Trials," New *England Journal of Medicine* 310:674-683, *1984.*
142. Park, R. E., Fink, A., Brook, R. H., et al., "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures," *American Journal of Public Health* 76:766-772, *1986.*
143. Pashos, C. L., Harvard University, Boston, MA, personal communication, August 1993.
144. Passamani, E., "Clinical Trials—Are They Ethical?" New *England Journal of Medicine* 324:1589-1592, *1991.*
145. Paul, J. E., Burroughs Wellcome, Research Triangle Park, NC, personal communications, August 1993.
146. Rahimtoola, S. H., "A Perspective on the Three Large Multicenter Randomized Clinical Trials of Coronary Bypass Surgery for Chronic Stable Angina," *Circulation 72* (suppl. V): V123-V135, 1985.
147. Ray, W.A., Griffin, M.R., "Use of Medicaid Data for Pharmacoepidemiology," *American Journal of Epidemiology* 129:837-849,1989.
148. Ray, W. A., Griffin, M. R., and Baugh, D. K., "Mortality Following Hip Fracture Before and After Implementation of the Prospective Payment *System, ''Archives of Internal Medicine* 150:2109-21 *14, 1990.*
149. Rimes, C., U.S. Department of Health and Human Services, Health Care Financing Administration, Office of National Health Statistics, Baltimore, MD, personal communication, September 1993.
150. Roberts, M. S., D'Agostino, R. B., Dillon, M., et al., "Technology Assessment in the Framingham Heart Study," *International Journal of Technology Assessment in Health Care 7: 156-170, 1991.*
151. Roethlisberger, F.J., and Dickson, W.J., *Management and the Worker* (Cambridge, MA: Harvard University Press, 1939).
152. Romano, P. S., and Luft, H. S., "Getting the Most of Out of the Messy Data: Problems and Approaches for Dealing with Large Administrative Data Bases," *Medical Effectiveness Research: Data Methods (Conference Proceedings, Summary Report,* M.L. Grady and H.A. Schwartz (eds.) (Rockville, MD: Agency for Health Care Policy and Research, 1992).
153. Roos, N. P., "Hysterectomy: Variations in Rates Across Small Areas and Across Physicians' Practice s," *American Journal of Public Health* 74:327-335,1984.
154. Roos, L.L., Cageorge, S. M., Austen, E., et al., "Using Computers To Identify Complications After Surgery," *American Journal of Public Health* 75: *1288-1295, 1985.*
155. Roos, N. P., and Lyttle, D., "Hip Arthroplasty Surgery in Manitoba: 1973 -1978," *Clinical Orthopedics and Related Research* 199:248-255, *1985.*
156. Roos, N. P., and Lyttle, D., "The Centralizations of Operations and Access to Treatment: Total Hip Replacement in Manitoba," *Amer-*

ican Journal of Public Health 75: *130-133, 1985.*

157. Roos, L. L., and Nicol, J. P., Johnson, C. F., et al., "Using Administrative Data Banks for Research and Evaluation: A Case Study," *Evaluation Quarterly* 3:236-255, 1979.

158. Roos, N.P., and Roos, L. L., "High and Low Surgical Rates: Risk Factors for Area Residents," *American Journal of Public Health* 71:591-600, *1981.*

159. Roos, L. L., Wajda, A., and Nicol, J. P., "The Art and Science of Record Linkage: Methods That Work with Few Identifiers," *Computers in Biology and Medicine* 16(1):45-57, *1986.*

160. Roos, L. L., Wajda, A., Nicol, J. P., Roberts, J., "Record Linkage: An Overview," *Medical Effectiveness Research: Data Methods,* M.L. Grady and H.A. Schwartz (eds) (Rockville, MD: Agency for Health Care Policy and Research, 1992).

161. Roos, N. P., Wennberg, J. E., Malenka, D.J., et al., "Mortality and Reoperation after Open and Transurethral Resection of the Prostate for Benign Prostatic Hyperplasia," *New England Journal of Medicine 320:* 1120-1124, 1989.

162. Roper, W. L., Winkenwerder, W., Hackbarth, G. M., et al., "Effectiveness in Health Care: An Initiative To Evaluate and Improve Medical Practice," New *England Journal of Medicine 319:* 1197-1202, 1988.

163. Rutkow, I. M., "Determining the Rates of Surgical Operations," *Socioeconomic of Surgery,* I.M. Rutkow (cd.) (St. Louis, MO: C.V. Moseby, 1989).

164. Sackett, D. L., "Bias in Analytic Research," *Journal of Chronic Disease* 32:51-63, 1979.

165. Sacks, H., "Randomized versus Historical Controls for Clinical Trials," *American Journal of Medicine* 72:233-240, *1982.*

166. Safran, C., "Using Routinely Collected Data for Clinical Research," *Statistics in Medicine* 10:559-564, *1991.*

167. SAS Institute Inc., *SAS User's Guide: Statistics: Version 5 Edition* (Cary, NC: SAS Institute Inc., 1985).

168. Savitz, D.A., and Grace, C., "Determinants of Medical Record Access for an Epidemiologic Study," *American Journal of Public Health* 75: *1425-1426,* 1985.

169. Sherman, C. R., Potosky, A. L., Weis, K. A., et al., "The Consensus Development Program. Detecting Changes in Medical Practice Following a Consensus Conference on the Treatment of Prostate Cancer," *International Journal of Technology Assessment in Health Care* 8:683-693, 1992.

170. Shipp, M. A., Yeap, B. Y., Barrington, D. P., et al, "The m-BACOD Combination Chemotherapy Regimen in Large-Cell Lymphoma: Analysis of the Completed Trial and Comparison with the M-BACOD Regimen," *Journal of Clinical Oncology* 8:84-93, 1990.

171. Simborg, D. W., "DRG Creep: A New Hospital-Acquired Disease," New *England Journal of Medicine 304: 1602-1604, 1981.*

172. Sloan, F. A., Morrisey, M. A., and Valvona, J., "Medicare Prospective Payment and the Use of Medical Technologies in Hospitals," *Medical Care* 26(9):837-853, *1988.*

173. Smits, H. L., and Watson, R. E., "DRGs and the Future of Surgical Practice," New *England Journal of Medicine 311:* 1612-1615, 1984.

174. Steinberg, E. P., John Hopkins Medical Institutions, Baltimore, MD, personal communication, May 1993.

175. Steinberg, E.P., Whittle, J., and Anderson, G. F., "Impact of Claims Data Research on Clinical Practice," *International Journal of Technology Assessment in Health Care* 6:282-287, *1990.*

176. Strom, B. L., Carson, J. L., Halpem, A. C., et al., "Using a Claims Database To Investigate Drug-Induced Stevens-Johnson Syndrome," *Statistics in Medicine* 10:565-576, *1991.*

177. Taylor, K. M., Margolese, R. G., and Soskolne, C. L., "Physicians' Reasons for Not Entering Eligible Patients in a Randomized Clinical Trial of Surgery for Breast Cancer," New *England Journal of Medicine 310:* 1363-*1367, 1984.*

178. Tierney, W. M., and McDonald, C.J., "Practice Databases and Their Uses in Clinical Research," *Statistics in Medicine* 10:541-557, *1991.*

179. Tunis, S. R., Bass, E. B., and Steinberg, E.P., "The Use of Angioplasty, Bypass Surgery, and Amputation in the Management of Peripheral Vascular Disease," New *England Journal of Medicine* 325:556-562,1991.

180. Udvarhelyi, 1. S., Gatsonis, C., Epstein, A. M., et al., "Acute Myocardial Infarction in the Medicare Population: Process of Care and Clinical Outcomes," *Journal of the American Medical Association 268:2530-2536, 1992.*

181. U.S. Department of Health and Human Services, *ICD-9-CM: The International Classification of Diseases, 9th Revision, Clinical Modification,* PHS Pub. No. 80-1260 (Washington, DC: U.S. Department of Health and Human Services, 1980).

182. U.S. Department of Health and Human Services, Health Care Financing Administration, Office of Statistics and Data Management, *Public Use Files Price List (July 1, 1993)* (Baltimore, MD: 1993).

183. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Conference Proceedings: Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data,* (Washington, DC: Agency for Health Care Policy and Research, 1990).

184. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Report to Congress: Progress of Research on Outcomes of Health Care Services and Procedures* (Rockville, MD: Public Health Service, 1991).

185. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research: *Report to Congress: The Feasibility of Linking Research-Related Data Bases to Federal and Non-Federal Medical Administrative Data Bases,* AHCPR Pub. No. 91-0003 (Rockville, MD: Public Health Service, 1991).

186. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Medical Electiveness Research Data Methods,* (Rockville, MD: Agency for Health Care Policy and Research, 1992).

187. U.S. General Accounting Office, *Cross Design Synthesis: A New Strategy for Medical Effectiveness Research* (Washington, DC: U.S. Government Printing Office, 1992).

188. U.S. General Accounting Office, *Cross Design Synthesis: A New Strategy for Medical Effectiveness Research* (Washington, DC: U.S. Government Printing Office, 1992).

189. Wajda, A., Roos, L. L., "Simplifying Record Linkage: Software and Strategy," *Computers in Biology and Medicine* 17(4):239-248, *1987.*

190. Wennberg, J. E., "Physician Uncertainty, Specialty Ideology, and a Second Opinion Prior to Tonsillectomy," *Pediatrics* 59:952, *1977.*

191. Wennberg, J. E., "On Patient Need, Equity, Supplier-Induced Demand, and the Need To Assess the Outcome of Common Medical Practices," *Medical Care* 23:512-520, *1985.*

192. Wennberg, J. E., "Population Illness Rates Do Not Explain Population Hospitalization Rates: A Comment on Mark Blumberg's Thesis That Morbidity Adjusters Are Needed To Interpret Small Area Variations," *Medical Care* 25:354-359,1987.

193. Wennberg, J. E., "The Paradox of Appropriate Care," *Journal of the American Medical Association* 258:2568-2569, *1987.*

194. Wennberg, J.E., Barnes, B. A., and Zubkoff, M., "Professional Uncertainty and the Problem of Supplier-Induced Demand," *Social Science in Medicine* 16:81 *1-824, 1982.*

195. Wennberg, J. E., Blowers, L., Parker, R., et al., "Changes in Tonsillectomy Rates Associated With Feedback and Review," *Pediatrics* 59:821-826, *1977.*

196. Wennberg, J., and Gittlesohn, "A., "Small Area Variations in Health Care Delivery," *Science* 182:1 *102-1108, 1973.*

197. Wennberg, J. E., Gittlesohn, A., and Soule, D., "Health Care Delivery in Maine II: Conditions Explaining Hospital Admission," *Journal of the Maine Medical Association* 66(10):255-261, *269, 1975.*

198. Wennberg, J.E., Mulley, A.G. Jr., Hanley, D., et al., "An Assessment of Prostatectomy for Benign Urinary Tract Obstruction: Geographic Variations and the Evaluations of Medical Care Outcomes," *Journal of the American Medical Association 259:3027-3030, 1988.*

199. Wennberg, J. E., Roos, N., Sola, L., et al., "Use of Claims Data Systems To Evaluate Health Care Outcomes: Mortality and Reoperation Following Prostatectomy," *Journal of the American Medical Associations* 257:933-936, *1987.*

200. Wenneker, M.B., and Epstein, A. M., "Racial Inequalities in the Use of Procedures for Patients with Ischemic Heart Disease in Massachusetts," *Journal of the American Medical Association* 261:253-257,1989.

201. Wenneker, M. B., Weissman, J. S., and Epstein, A. M., "The Association of Payer with Utilization of Cardiac Procedures in Massachusetts," *Journal of the American Medical Association* 264:1255-1260, *1990.*

202. Whittle, J., Conigliaro, J., Good, C. B., et al., "Racial Differences in the Use of Invasive Cardiovascular Procedures in the Department of Veterans Affairs Medical System," New *England Journal of Medicine* 329:621-627, *1993.*

203. Whittle, J., Steinberg, E. P., Anderson, G. F., et al., "Use of Medicare Claims Data To Evaluate Outcomes in Elderly Patients Undergoing Lung Resection for Lung Cancer," *Chest* 100:729-734, 1991.

204. Whittle, J., Steinberg, E. P., Anderson, G. F., et al, "Mortality After Elective Total Hip Arthroplasty in Elderly Americans: Age, Gender, and Indication for Surgery Predict Survival," *Clinical Orthopedics and Related Research 295:* 119-126, 1993.

205. Winslow, C. M., Solomon, D. H., Chassin, M.R., et al., "The Appropriateness of Carotid Endarterectomy," New *England Journal of Medicine* 318:721-727, *1988.*

206. Wittes, R.E., "Paying for Patient Care in Treatment Research-Who Is Responsible?" *Cancer Treatment Reports* 71:107-1*13, 1987.*

207. Wyse, D.G., Kellen, J., and Rademaker, A. W., "Prophylactic versus Selective Lidocaine for Early Ventricular Arrhythmias of Myocardial Infarction, ''Journal *of the American College of Cardiology* 12:507-513,1988.

208. Wyshak, G., Burdick, E., and Mosteller, F., "Technology Assessment in the Connecticut Tumor Registry," *International Journal of Technology Assessment in Health Care* 7:129-133, *1991.*

209. Yancik, R., Ries, L.G., and Yates, J. W., "Breast Cancer in Aging Women: A Population-Based Study of Contrasts in Stage, Surgery and Survival," *Cancer* 63:976-981, *1989.*

210. Young, J.L. Jr., Percy, C. L., and Asire, A.J., *Surveillance, Epidemiology, and End Results: Incidence and Mortality Data, 1973-1977* (Washington, DC, National Cancer Institute, 1981).