Looking for Answers in All the Wrong Places

An old vaudeville story describes a man groping on all fours beneath a lamppost looking for a quarter he had dropped a block away. When asked his reason for searching there, he responded, "because the light's better." The article by Jollis and colleagues (1) in this issue of *Annals* brings this story to mind. They and other members of the cardiology division at Duke University have for years prospectively accumulated a rich clinical database for managing patients with cardiovascular disease. Consequently, they were in a unique position to answer the Agency for Health Care Policy and Research's call of PORT (Patient Outcome Research Team) initiative to use the massive Medicare billing database to answer questions about effectiveness and outcome for patients with ischemic heart disease (2). Their study shows, not surprisingly, that data sets completed for billing purposes and constructed mainly by financial experts differ substantially from those constructed by clinicians caring for patients.

Previous retrospective studies of the accuracy of claims data for the diagnosis of acute myocardial infarction found that clinical criteria were met in 43% to 80% of patients discharged with that diagnosis (3–6). The Duke study extends these findings because it did not use retrospective chart review to validate the diagnosis but compared claims data with contemporaneously collected clinical data. In addition, the investigators were not constrained in identifying patients on the basis of claims data alone.

Their study involved 12 937 consecutive patients discharged between July 1985 and May 1990 with a procedure code for coronary arteriography. Lacking a third comparison group, they used the cardiology data set as the criterion or "gold" standard because it was likely to be more accurate. Although the assumption probably is correct, purists might give the claims data the benefit of the doubt by assigning accuracy whenever a diagnosis appears in either data set. Recalculating their results using the latter assumption (Table 1) does not substantively change their findings. The strongest agreement between the two data sets was in the diagnosis of diabetes: This diagnosis was missing in 15% of the cases in the claims data set and in 9% of the cases in the cardiology data set. For acute myocardial infarction, the discrepancy was 23% and 7%, respectively. The discrepancy for important prognostic variables for ischemic heart disease was even greater. The diagnosis of congestive heart failure was missing in 52% of the cases in the claims data set but in only 19% of the cases in the cardiology data set. For the diagnosis of mitral insufficiency, the discrepancy was 49% and 12%, respectively; for the diagnosis of unstable angina, the discrepancy was 69% and 3%. Overall, more than half the patients with prognostically important conditions were not identified by the claims data set.

Although discrepancies in the coding and documentation of billing data are not new, they were of little concern to most physicians until they began to affect reimbursement and judgments about quality of care. In 1982, Maryland's Health Care Cost Review Commission uniform data sets for hospitals were first used to see if billed diagnoses justified permanent pacemaker insertion. Ralph Nader's Public Citizen Health Research Group concluded that at least 817 of 2222 insertions were unjustified (7). A re-review of 610 of these questionable 817 placements showed, in 95% of cases, failure to code justifying diagnoses or erroneous procedure coding (8). A follow-up to a companion report assigning surgical mortality rates to Maryland hospitals (9, 10) also showed many salient coding errors, including instances in which patients were inaccurately reported as having been discharged dead. Since then, the incentives to pay attention to coding have increased, and, as Jollis and colleagues show, the claims data sets have improved. Between 1985 and 1990, the likelihood of identifying important clinical conditions increased from 33% to 46% — good but not great. Much of the improvement was for patients older than 64 years, probably because of Medicare requirements.

The temptation to use databases for research because they are large and include a defined population such as the elderly is understandable (11, 12). Dissemination of increasingly sophisticated, user-friendly computer hardware and software now permits the manipulation of gigantic databases, leading many to say that we are in the midst of an "information explosion." However, it is

Table 1. Concordance of Cardiology and Claims Databases

Condition	Number of Patients Identified Using Either Data Set	Cardiology Data Set*	Claims Data Set†
		%	
Diabetes mellitus	2727	91	85
Acute myocardial infarction	5398	93	77
Hypertension	7109	92	68
Mitral insufficiency	2753	88	51
Congestive heart failure	2209	81	48
Peripheral vascular disease	1519	81	42
Old myocardial infarction	3315	91	36
Hyperlipidemia	4685	84	47
Cerebrovascular disease	1139	88	24
Tobacco abuse	8298	99	25
Angina	10001	97	31
Unstable angina	7680	97	16

* Number of "correct" diagnoses in cardiology data set over number in both sets.

† Number of "correct" diagnoses in claims data set over the number in both sets.

more appropriately called a "data explosion" because databases, no matter how large or how available, are not informative unless they are accurate, complete, and can answer important questions. Otherwise, they are distracting and potentially misleading.

More subtle are the undue effects on young investigators, who, because of the pressure to publish, often seek out established databases to mine. The approach seems to be: identify the data first, then proceed to the question. There is even software that promises to find the buried treasure in a database. This approach is the reverse of good science and at its worst can lead to data dredging and misleading posthoc analyses, not because the questions are unimportant but because the database is not up to the task. Indeed, complex statistical methods may lend false credibility to predictive models constructed from under-representative data or unvalidated adjustments for key missing variables.

The challenge, then, is to use large data sets appropriately. The Medicare data set, for example, is useful for analyzing trends in efficiency and charges using such process measures as paid claims, length of stay, and resource use. As for outcomes, when linked to the Social Security files, the Medicare data set can help to identify trends in mortality and to generate explanatory hypotheses (13). However, to refute or confirm these hypotheses requires specific prospectively collected clinical information that is not readily available.

One major reason people turn to such databases is that their availability makes using them comparatively inexpensive (11, 12), whereas conducting the studies to gather primary data on how patients are managed is costly. In 1977, Congressman Henry Waxman introduced a bill (HR 4869) to create a new institute at the National Institutes of Health (NIH) devoted to studies of cost-effectiveness and technology assessment (14). He believed that placement within NIH would enhance the prestige of these neglected areas among clinical researchers and at the same time ensure its adequate funding because of the NIH's political standing in Congress. The NIH opposed the bill largely because such applied research was so potentially costly that it might overshadow their basic research mission; instead, NIH created an Office of Medical Applications and Research to convene consensus conferences on important questions but not to produce new data (15). Simultaneously, Senator Edward Kennedy successfully introduced legislation that passed in 1978 as Public Law 95-623 creating a National Center for Health Care Technology, but the Center survived only 3 years because of political and institutional problems (16). Since then, the Agency for Health Care Policy Research, established in 1989 (17), has funded studies to determine outcomes in such diverse conditions as prostate cancer, back pain, biliary tract disease, cataracts, and hip fracture. However, allocation for research and development to assure quality outcomes and cost-effective care is probably less than $150 million, or about 0.02% of what is spent on health care. Any meaningful health care reform must substantially increase this percentage.

Outcome studies will require that physicians enter key diagnostic and prognostic data in an accurate and timely manner. This is more difficult than it sounds, not just because of the enormous logistical problems or the need for positive incentives to assure completeness and to inhibit overcoding, but also because of the lack of physician uniformity in naming patient conditions. For example, "unstable angina," an important prognostic indicator, is variably used by both cardiologists and noncardiologists. One doctor's "bronchitis" may be another's "bronchopneumonia." For the most part, these discrepancies reflect differences in physician training and temperament. All of us can recall the range in the admitting thresholds of residents, from low ("sieves") to extremely high ("iron-gates"). Given that, in Osler's words (18), "medicine is a science of uncertainty and an art of probability," it is not surprising that different thought patterns may lead to different diagnostic and treatment strategies. As Chassin (19) has suggested, much of the demonstrated small area variation in medical care may reflect the enthusiasm of local clinician opinion makers, abetted by overly optimistic reporting of research results and financial incentives or patient and family pressures. Such enthusiasm can also be seen as a physician response to Feinstein's "chagrin factor" by using all available resources to minimize regret (20).

Thus, the mandate to link databases (21, 22) and to collect data that are more accurate and complete (23) should be seen not as a means of necessarily yielding "the answer" but of ushering in a much-needed explosion in research on how, why, and to what end physicians do what they do. Indeed, in my experience, good physicians want to know not just "How am I doing?" but "How can I do better?" Unfortunately, clinicians have not been sufficiently involved in framing the important questions and in ensuring that the necessary data are collected. This has generated legitimate fears that extant data sets can be misused by persons who are not themselves accountable to patients to make definitive and too often pejorative judgments rather than to generate hypotheses and to educate physicians and patients. They are right to be concerned. Benefits managers and insurers are now faced with a proliferation of physician profiling programs that rely on ICD-9-CM–disease coding to aggregate patients and to adjust for severity of illness using techniques that have not been widely validated. Researchers involved in such efforts need to be careful, then, not just to use the right data sets but to specify what their studies really show. Although investigators may state the limitations of their data in the discussion section of their papers, such qualifiers do not usually appear in the television and press accounts. In short, as we enter the outcome study era (24), health care policymakers and researchers need to avoid, to paraphrase a 1980 country song, "Looking for answers in all the wrong places."

Peter E. Dans, MD
Deputy Editor
Annals of Internal Medicine
Philadelphia, PA 19106

Requests for Reprints: Peter E. Dans, MD, American College of Physicians, Independence Mall West, Sixth Street at Race, Philadelphia, PA 19106-1572.

Ann Intern Med. 1993;119:855-857.

References

1. **Jollis JG, Ancukiewicz M, Delong ER, Pryor DB, Muhlbaier LH, Mark DB.** Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. Ann Intern Med. 1993;119:844-50.
2. **Sallve ME, Mayfield JA, Weissman NW.** Patient outcomes research teams and the Agency for Health Care Policy and Research. Health Serv Res. 1990;25:697-708.
3. **Kennedy GT, Stern MP, Crawford MH.** Miscoding of hospital discharges as acute myocardial infarction: implications for surveillance programs aimed at elucidating trends in coronary artery disease. Am J Cardiol. 1984;53:1000-2.
4. **Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA.** Coding of acute myocardial infarction. Clinical and policy implications. Ann Intern Med. 1988;109:745-51.
5. **Shiff GD, Yaacoub AS.** The diagnostic coding of myocardial infarction [Letter]. Ann Intern Med. 1989;110:243.
6. **Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al.** The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992; 82:243-8.
7. **Greenberg A, Kowey PR, Bargmann E, Wolfe SM.** Permanent pacemakers in Maryland. Washington, D.C.: Health Research Group; 1982.
8. **Scherlis L, Dembo DH.** Problems in health data analysis: the Maryland permanent pacemaker experience in 1979 and 1980. Am J Cardiol. 1983;51:131-6.
9. **Burgmann E, Grove C.** Surgery in Maryland Hospitals, 1979 and 1980: Charges and Deaths. Washington, D.C.: Health Research Group; 1982.
10. **Zuidema GD, Dans PE, Dunlap ED.** Documentation of care and prospective payment. One hospital's experience. Ann Surg. 1984; 199:515-21.
11. **Steinberg EP, Whittle J, Anderson GF.** Impact of claims data research on clinical practice. Int J Technol Assess Health Care. 1990;6:282-7.
12. **Romano P, Luft HS.** Getting the most out of messy data: problems and approaches for dealing with large administrative data sets. In: Medical Effectiveness Research Data Methods. Rockville, Maryland: U.S. Department of Health and Human Services; 1992:57-75. AHCPR pub. no. 92-0056.
13. **Whittle J, Steinberg EP, Anderson GF, Herbert R.** Use of Medicare claims data to evaluate outcomes in elderly patients undergoing lung resection for lung cancer. Chest. 1991;100:729-34.
14. **Dans PE.** Issues along the Potomac: "efficacy and technology transfer." South Med J. 1977;70:1225-31.
15. **Perry S, Kalberer JT Jr.** The NIH consensus development program and the assessment of health-care technologies: the first 2 years. N Engl J Med. 1980;303:169-72.
16. **Blumenthal D.** Federal policy toward health care technology: the case of the National Center. Milbank Q. 1983;61:584-613.
17. **Fitzmaurice JM.** What do we do when the data for medical effectiveness research are not perfect? In: Medical Effectiveness Research Data Methods. Rockville, Maryland: U.S. Department of Health and Human Services; 1992:5-7.
18. **Bean WB, ed.** Sir William Osler: Aphorisms. From His Bedside Teachings and Writings. Springfield, Illinois: Thomas; 1961.
19. **Chassin MR.** Explaining geographic variations: The enthusiasm hypothesis. Med Care. 1993;31(5 Suppl):YS37-44.
20. **Feinstein AR.** The "chagrin factor" and qualitative decision analysis. Arch Intern Med. 1988;145:1257-9.
21. The Feasibility of Linking Research-related Data Bases to Federal and Non-federal Medical Administrative Data Bases. AHCPR Pub. No. 91-0003. Rockville, Maryland: U.S. Department of Health and Human Services; 1991.
22. **Automated Data Sources for Ambulatory Care Effectiveness Research.** AHCPR Pub. No. 93-0042. Rockville, Maryland: U.S. Department of Health and Human Services; 1993.
23. **Helthoff KA, Lohr KN.** Effectiveness and Outcomes in Health Care: Proceedings of an Invitational Conference by the Institute of Medicine, Division of Health Care Services. Washington, D.C.: National Academy Press; 1990.
24. **Wennberg JE.** Outcomes research, cost containment, and the fear of health care rationing. N Engl J Med. 1990;323:1202-4.