Annals of Internal Medicine

Assessing Quality Using Administrative Data

Lisa I. Iezzoni, MD, MSc

Administrative data result from administering health care delivery, enrolling members into health insurance plans, and reimbursing for services. The primary producers of administrative data are the federal government, state governments, and private health care insurers. Although the clinical content of administrative data includes only the demographic characteristics and diagnoses of patients and codes for procedures, these data are often used to evaluate the quality of health care. Administrative data are readily available, are inexpensive to acquire, are computer readable, and typically encompass large populations. They have identified startling practice variations across small geographic areas and supported research about outcomes of care. Many hospital report cards (which compare patient mortality rates) and physician profiles (which compare resource consumption) are derived from administrative data. However, gaps in clinical information and the billing context compromise the ability to derive valid quality appraisals from administrative data. With some exceptions, administrative data allow limited insight into the quality of processes of care, errors of omission or commission, and the appropriateness of care. In addition, questions about the accuracy and completeness of administrative data abound. Current administrative data are probably most useful as screening tools that highlight areas in which quality should be investigated in greater depth. The growing availability of electronic clinical information will change the nature of administrative data in the future, enhancing opportunities for quality measurement.

State and regional efforts to assess the quality of health care often start with administrative data, which are a by-product of administering health services, enrolling members into health insurance plans, and reimbursing for health care services. By definition, administrative data were never intended for use in quality assessment. As a result, clinicians often dismiss these data, arguing that the information cannot be trusted. Nonetheless, with detailed clinical information buried deep within paper medical records and thus expensive to extract, adminis-

trative data possess important virtues. They are readily available; are inexpensive to acquire; are computer readable; and typically encompass entire regional populations or large, well-defined subpopulations.

In the health policy community, hopes for administrative data were initially high. Beginning in the early 1970s, administrative data quantified startling practice variations across small geographic areas (1, 2). In the 1980s, administrative databases became a mainstay of research on the outcomes of care (3, 4). In 1989, legislation that created the Agency for Health Care Policy and Research (AHCPR) stipulated the use of "claims data ... in determining the outcomes, effectiveness, and appropriateness" of different therapies (Public Law 101-239, Section 1142(c)). Five years later, however, the Office of Technology Assessment offered a stinging appraisal: "Contrary to the expectations expressed in the legislation establishing AHCPR ... administrative databases generally have not proved useful in answering questions about the comparative effectiveness of alternative medical treatments" (5).

The costs of acquiring detailed clinical information, however, often force concessions in the "real world." For example, in 1990, California's Assembly debated new requirements for reporting clinical data to evaluate hospital quality (6). When estimated annual costs for data collection were $61 million, fiscal reality intervened. The legislature mandated the creation of quality measures that used California's existing administrative database.

Thus, widespread quality assessment typically demands a tradeoff—the credibility of clinical data versus the expense and feasibility of data collection. Can administrative data produce useful judgments about the quality of health care?

Defining Quality

What is quality? For decades, physicians protested that defining health care quality was impos-

sible. Today, however, experts claim that rigorous quality measures can systematically assess care across groups of patients (7, 8). Nonetheless, consensus about specific methods for measuring quality remains elusive. Different conceptual frameworks for defining quality stress different dimensions of health care delivery.

Donabedian's classic framework (9) delineated three dimensions: 1) structure, or the characteristics of a health care setting (for example, the physical plant, available technology, staffing patterns, and credentialing procedures); 2) process, or what is done to patients; and 3) outcomes, or how patients do after health care interventions. The three dimensions are intertwined, but their relative utility depends on context. Few links between processes and outcomes are backed by solid evidence from well-controlled studies, and outcomes that are not linked to specific medical practices provide little guidance for developing quality-improvement strategies (10). In addition, comparing outcomes across groups frequently requires adjustment for patient risk and the recognition that some patients are sicker than others (11). Other important dimensions emerge when a process splits into two components: technical quality and interpersonal quality (for example, communication, caring, and respect for patient preferences). Another process question involves the appropriateness of services: errors of omission (failing to do necessary things) and errors of commission (doing unnecessary things). Both errors can be related to another important dimension of quality: access to health care. In errors of omission, access may be impeded; in errors of commission, access may be too easy or inducements to perform procedures too great.

In today's environment, determining who (or what) is accountable for observed quality is as important as measuring quality. This requires defining a unit of analysis: quality for whom? Potential units of analysis include individual patients, patients grouped by providers, or populations defined by region or an important characteristic (for example, the insurer or patient age). Methods for measuring quality across populations differ from those that scrutinize quality for individual patients.

Given these multidimensional perspectives, a single response may be insufficient to judge whether administrative data can assess health care quality. As discussed in the following sections, administrative data may capture some dimensions of quality and units of observation better than others.

Content of Administrative Databases

The three major producers of administrative databases are the federal government (including the

Table 1. Contents of the Uniform Hospital Discharge Data Set

Personal identification
Date of birth
Sex
Race and ethnicity
Residential zip code
Hospital identification
Admission date
Discharge date
Attending physician identification
Operating physician identification
Codes for principal diagnosis and other diagnoses
Codes and dates for principal procedure and other procedures
Disposition of the patient
Expected principal source of payment

Health Care Financing Administration [HCFA], which administers Medicare and oversees Medicaid; the Department of Defense; and the Department of Veterans Affairs), state governments, and private insurers (3, 4, 12–19). Although administrative files initially concentrated on information from acute care hospitals, information is increasingly compiled from outpatient, long-term care, home health, and hospice programs. Most administrative files explicitly aim to minimize data collection. Their source documents (for example, claim forms) contain the minimum amount of information required to perform the relevant administrative function (for example, to verify and pay the claims). In this article, I focus on hospital-derived data (such as that obtained from discharge abstracts), but many of the issues examined apply to other care settings.

Their clinical content delimits the potential of databases to measure the quality of health care. Administrative sources always contain routine demographic data (**Table 1**). Additional clinical information includes diagnosis codes (based on the *International Classification of Diseases, Ninth Revision, Clinical Modification* [ICD-9-CM]) and procedure codes. Hospitals report procedures using the ICD-9-CM codes, but physicians generally use codes from the American Medical Association's *Current Procedural Terminology*. The two coding systems do not readily link, hindering comparisons between hospital- and physician-generated data.

The ICD-9-CM contains codes for many conditions that are technically not diseases (**Table 2**). Given this diversity, creatively combining ICD-9-CM codes produces snapshots of clinical scenarios. For example, data selected from the 1994 discharge abstract of a man in a California hospital (**Table 3**) suggest the following scenario: A 62-year-old white man with a history of chronic renal failure that required hemodialysis and type 2 diabetes with retinopathy was admitted with the Mallory-Weis syndrome. Blood loss from an esophageal tear may have caused orthostatic hypotension. During the 9-day

hospitalization, the patient was also treated for *Klebsiella* pneumonia.

This diversity of ICD-9-CM codes is used by administrative data–based severity measures (20–22) aiming to compare risk-adjusted patient outcomes across hospitals. For example, Disease Staging rates patients with pneumonia as having more severe disease if the discharge abstract also contains codes for sepsis.

Attributes of Administrative Data

Administrative files contain limited clinical insight to inform quality assessment. Administrative data cannot elucidate the interpersonal quality of care, evaluate the technical quality of processes of care, determine most errors of omission or commission, or assess the appropriateness of care.

Some exceptions to these negative judgments do exist. For example, with longitudinal person-level data, one could detect failures to immunize children (errors of omission)—if all immunizations were coded properly, which is unlikely. Certain ICD-9-CM procedure codes prompt concerns about technical quality (for example, 39.41, control of hemorrhage after vascular surgery, and 54.12, reopening of recent laparotomy site), but the specificity of the codes is suspect. Nonetheless, administrative data are widely used to produce hospital report cards that primarily compare in-hospital mortality rates.

The mechanics are easy. For example, in Massa-

Table 2. Examples of Information Contained in ICD-9-CM Codes*

Information	Code	Code Name
Clinical diagnosis	491.0	Simple chronic bronchitis
Extent of clinical diagnosis	250.43	Type 1 diabetes mellitus with renal manifestations, uncontrolled
Pathologic process	580.4	Acute glomerulonephritis with lesion of rapidly progressive glomerulonephritis
Symptoms	784.0	Headache
Physical findings	611.72	Lump or mass in breast
Laboratory or other test findings	790.3	Excessive level of blood alcohol
Severity indicators	427.5	Cardiac arrest
Potential quality indicators	998.2	Accidental puncture or laceration during a procedure
Psychological factors	V15.81	Personal history of noncompliance with medical treatment
Cognitive factors	318.0	Moderate mental retardation (IQ 35–49)
Substance abuse	304.21	Continuous cocaine dependency
Personal and social factors	V60.0	Homelessness
Functional status	344.1	Paraplegia
External environmental factors	E900.0	Excessive heat due to weather conditions

* ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification.

Table 3. Discharge Abstract Information for a Patient Admitted to a California Hospital in 1994*

Demographic information
  Age: 62 years
  Sex: Male
  Race: White
Administrative information
  Length of stay: 9 days
  Charges: $38 753
  Payor: Self pay
  Disposition: Routine discharge
  Admission type: Urgent
  DRG 174: Gastrointestinal hemorrhage with complication/comorbidity
ICD-9-CM diagnosis codes
  530.7 Gastroesophageal laceration–hemorrhage syndrome
  250.50 Type 2 diabetes mellitus with ophthalmic manifestations, not stated as uncontrolled
  362.01 Background diabetic retinopathy
  585 Chronic renal failure
  482.0 Pneumonia due to *Klebsiella pneumoniae*
  787.0 Nausea and vomiting
  458.0 Orthostatic hypotension
  V45.1 Renal dialysis status
ICD-9-CM procedure codes
  45.13 Esophagogastroduodenoscopy
  54.98 Peritoneal dialysis
  39.95 Hemodialysis
  88.01 CAT scan of abdomen
  93.39 Other physical therapy
  88.76 Diagnostic ultrasound of abdomen and retroperitoneum

* CAT = computerized axial tomography; DRG = diagnosis-related group; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification.

chusetts, reporters for *The Boston Globe* purchased the state's database of hospital discharge abstracts, conducted analyses, and published a report card on hospital mortality. The report card was explicitly intended to provide insight into the quality of health care (23). Are quality assessments based on administrative data valid? As Donabedian observed (9), a major aspect of validity "has to do with the accuracy of the data." The Institute of Medicine's Committee on Regional Health Data Networks made the reliability and validity of data an absolute requirement that had to be satisfied before public dissemination of derived quality measures (12):

> The public interest is materially served when society is given as much information on costs, quality, and value for health care dollar expended as can be given accurately.... Public disclosure is acceptable *only* when it: (1) involves information and analytic results that come from studies that have been well conducted, (2) is based on data that can be shown to be reliable and valid for the purposes intended, and (3) is accompanied by appropriate educational material.

What, therefore, are the important attributes of administrative data?

Data Quality

Like quality of care, quality of data encompasses several dimensions.

Clinical Content

An initial question is whether ICD-9-CM diagnosis codes accurately and completely represent clinical presentations (24). Although the ICD-9-CM sys-

tem includes approximately 15 000 diagnosis codes, it does not provide clinical descriptions to define each code. For example, it lists 39 four- and five-digit codes for different types of anemia but does not specify the hematocrit that justifies these diagnoses. In addition, despite the large number of codes, ICD-9-CM does not include many characteristics with significant prognostic indicators. The codes for anemia merely indicate that anemia was present; they do not specify the actual hematocrits or how rapidly they developed. The ICD-9-CM system is unable to capture many clinical problems that are typically encountered in outpatient settings (25) and important functional, socioeconomic, and psychosocial factors.

Coding Accuracy

The accuracy of diagnosis coding obviously affects data quality. The motivation of diagnostic coding changed with the 1983 enactment of Medicare's prospective payment system, which is based on diagnosis-related groups (DRGs). Because of code creep ("a deliberate and systematic shift in a hospital's reported case mix in order to improve reimbursement" [26]), the accuracy of coding for diagnoses became suspect. Such words as *optimization* and *maximization* entered the coding vocabulary. Some hospitals shifted medical record departments from general administration to financial divisions (27).

Nationwide studies of coding accuracy were performed by Medicare's legal overseer, the Office of the Inspector General. The first study sampled Medicare admissions between October 1984 and March 1985 in 239 hospitals (28) and found that 61.7% of the 20.8% identified coding discrepancies financially favored the hospitals (that is, hospitals had assigned DRGs that had higher relative weights). As a result, the government required physicians to attest to the accuracy of diagnostic codes before hospitals submitted invoices for Medicare payment. The Office of the Inspector General repeated the study by using records from 1988 (29). Results showed that 14.7% of records contained errors that changed the assignment of diagnosis-related groups; 50.7% of the errors financially benefited hospitals. Code creep appeared cured. However, in September 1995, Vice President Gore eliminated the attestation form to simplify the paperwork required by federal health programs (30, 31), leading to concern about the recrudescence of code creep. In addition, as health care is increasingly provided on an outpatient basis, Medicare's concerns about the accuracy of outpatient coding are mounting (32).

Coding accuracy became a political liability for California's mandated hospital report cards derived from administrative data (33). Responding to a maelstrom of criticism, the Office of Statewide Health Planning and Development abstracted 974 heart attack records and found substantial variations in the accuracy of coding for some risk factors (for example, hypotension, pulmonary edema, nutritional deficiency, chronic liver disease, and late effects of stroke). California no longer uses these poorly coded conditions in risk-adjustment of mortality rates.

Completeness of Coding

Administrative data are typically submitted in formats that limit the number of coding slots. For example, the old billing form used by HCFA and many current state discharge abstracts allow only five diagnosis codes and three procedure codes. Although five diagnoses may be sufficient for uncomplicated admissions, this number is often inadequate for complicated admissions or patients with multiple conditions. A study of Medicare data (34) suggested that chronic conditions were less likely to be coded when patients died because all the coding slots were consumed by acute diagnoses. This possibility prompted HCFA to increase the number of coding slots. However, studies from California (which has long offered 25 coding slots for diagnosis and 25 slots for procedure) suggest that increasing the number of diagnosis slots may not improve the completeness of coding for chronic conditions (35, 36).

Differences in Data Quality across Hospitals

Variability in data quality across hospitals compromises the utility of administrative data for comparing hospital performance. Because hospitals code with different degrees of thoroughness and accuracy, one cannot tell whether coding differences reflect true differences between patients without further study. The reabstraction study from California of 974 patients with heart attacks found variations in coding accuracy across hospitals. Overall, at least one clinical risk factor was missing for 65.0% of patients (range, 45% to 87% across hospitals) (33). In contrast, 31.5% of records contained at least one unsupported risk factor; this overcoding was more common at "low" mortality hospitals than at "high" mortality hospitals (36.7% compared with 29.0%). Overcoding ranged from 10% at one "high" mortality hospital to 74% at a "low" mortality hospital. Variation in coding accuracy explained part of the differences between "high" and "low" mortality hospitals (33).

Timing of Events

Knowing when events happen is crucial for assessing quality. However, the information in hospital discharge abstracts is retrospective: Diagnoses are

Table 4. c Statistics for Predicting In-Hospital Mortality*

Severity Measures and Data Source	Condition or Procedure			
	AMI (n = 11 880)	CABG Surgery (n = 7765)	Pneumonia (n = 18 016)	Stroke (n = 9407)
Clinical data–based measure				
MedisGroups	0.83	0.74	0.85	0.87
Discharge abstract–based measures				
All-Patient Refined and Diagnosis-Related Groups	0.84	0.83	0.78	0.77
Disease Staging mortality probability	0.86	0.78	0.80	0.74
Patient Management Categories severity score	0.82	0.81	0.79	0.73

* AMI = acute myocardial infarction; CABG = coronary artery bypass graft.

assigned after discharge. Discharge diagnoses reflect conditions that were diagnosed or treated at any time during the entire admission, regardless of when they occurred. Differentiating the timing of each diagnosis is key to risk adjustment; intrinsic risk factors must be separated from factors caused by substandard inpatient care. In the example given in **Table 3**, chronic renal failure and type 2 diabetes are obviously preexisting conditions. However, we cannot determine whether *Klebsiella* pneumonia was also preexisting or whether it was a nosocomial infection. (California has since added a flag to indicate whether a diagnosis was present on admission. New York is the only other state that indicates the timing of diagnosis.)

As predictors of in-hospital mortality, discharge abstract–based severity measures are sometimes equal to or better than measures derived from clinical findings at admission (37–41) (**Table 4**). For example, for acute myocardial infarction, the discharge abstract–based All-Patient Refined and Diagnosis-Related Groups (22) and Disease Staging (20) produced slightly higher c statistics than did the clinical data–based MedisGroups derived from findings from the first 2 hospital days (42). Discharge abstract–based severity measures may perform well in predicting in-hospital mortality because they rely on ICD-9-CM codes for life-threatening conditions (such as cardiac arrest), regardless of when these conditions occurred (38, 43, 44).

One way to address this problem is to risk-adjust by using only ICD-9-CM codes for diagnoses that are unlikely to arise de novo during hospitalization, such as diabetes and chronic renal failure (33, 43, 45). In addition, longitudinal data could identify conditions that had been treated previously and would thus be considered chronic or preexisting. For the patient in **Table 3**, an outpatient visit for pneumonia (even on the day of admission) would provide important insight into the timing of events.

Structural Attributes

Structural aspects of administrative databases could affect quality measurement, especially in population-based studies. The ability to track services used by persons across care settings (such as hospitals, private physicians' offices, and nursing homes) enhances the power of administrative data. Providing the necessary linkage, however, typically requires the use of unique patient identification numbers (regardless of payer or provider), which arouses fears about confidentiality and privacy. Databases from Medicare, Veterans Affairs, Medicaid, and a few states contain unique patient identifiers. Nevertheless, even without these identifiers, records can still be linked with reasonable success if sufficient demographic and administrative data are available (46). Some databases (for example, the Medicare database) also link readily eligibility records that indicate dates of death.

Another crucial structural question is whether denominators (that is, populations covered) can be determined so that rates of events can be calculated. Denominators facilitate population-level quality assessment, assuming that meaningful events (numerators) are specified. Enrollment files maintained by Medicare are detailed and updated frequently; in contrast, eligibility for Medicaid fluctuates over time for many persons. Crossing borders to obtain services and residential migration from one state to another pose substantial problems for some state databases. Determining populations encompassed by private insurance claims files can be particularly daunting, given that many insurers update enrollment files only when groups renew coverage (19).

Administrative Data and Quality Measurement

Despite the concerns discussed above, it is possible to patch together potential indicators of quality from administrative data, primarily around three dimensions of quality: access (for example, whether older women have mammography) (47), limited outcomes (for example, death or certain complications) (33), and limited processes (for example, whether

patients with diabetes receive retinal examinations) (48).

Even such modest administrative data–based quality standards may not, however, pass the first hurdle of validity: accuracy. Gaps in clinical information, questions about coding procedures, and the billing context restrict the ability to derive valid appraisals of quality from administrative data. The assessment of validity, however, is not a yes-or-no proposition: Shades of gray are possible. Even the Office of Technology Assessment has acknowledged that administrative data may be "very useful for descriptive purposes" (5). For quality assessment, an intermediate position holds that administrative data are a useful screening tool that highlights areas in which quality should be investigated in greater depth by using detailed clinical information.

In the practical world, screening is attractive. In-depth evaluations are costly, and methods that target scarce resources are valuable. In addition, as with California's report card on heart attack mortality rates (33), extensive technical documentation can detail caveats. This approach satisfies the stipulation of the Committee on Regional Health Data Networks that appropriate educational material accompany publicly disseminated quality measures (12).

Nonetheless, "real-world" considerations also prompt serious reservations about screening. As Thomas observed (49): "The unfortunate thing is that after these figures are published in the newspaper, all of the cautions ... in the original document are usually cast aside and the numbers are taken as facts." Hospitals could be unfairly flagged as "bad" because of data artifacts. Hannan and colleagues (50) used two data sources to examine in-hospital mortality for coronary artery bypass graft surgery: a clinical database produced by the Cardiac Surgery Reporting System and an administrative file from the New York Statewide Planning and Research Cooperative Systems. The Cardiac Surgery Reporting System contained detailed risk factors, including left ventricular ejection fraction, percent narrowing of the left main coronary artery, unstable angina, and previous myocardial infarction. The two sources disagreed about whether some hospitals were high or low mortality outliers, prompting the following observation (50):

> ... It can be argued that the importance of using a clinical data base rather than an administrative data base for CABG surgery is dependent on the purposes of the data base. If the information is used for internal quality assurance purposes or to target hospitals for state or federal site visits, the administrative data base may suffice. However, if the information is used to inform consumers of relative quality of hospital care, the differences in hospital ratings between the two systems as well as the potential damage to a hospital's reputation are probably too great to risk using an administrative data base.

Some complaints from hospitals may not elicit sympathy. When *The Boston Globe* identified 10 Massachusetts hospitals as "mortality outliers" (23), certain hospitals argued that the administrative data—produced by the hospitals themselves—had not contained all of the ICD-9-CM codes indicating how sick their patients were. This was true; reevaluation of the records showed that the hospitals could have reasonably assigned additional codes that would have better identified high-risk patients. When the hospitals originally coded these charts, they did not anticipate that the data would be used to risk-adjust mortality rates. Such arguments are difficult to convey in compelling "sound bytes."

"Death creep" is also possible, reprising the "code creep" epidemic in the mid-1980s; that is, coding may be used to improve mortality rates. If hospitals know that their discharge abstracts will be used to evaluate mortality rates, coding of near-death conditions (such as cardiac arrest, respiratory failure, or shock) for patients who die will increase. Assigning such codes is clinically reasonable (by definition, patients who die experience cardiac arrest), but such coding shifts would confuse efforts to risk-adjust mortality rates for useful comparisons across hospitals. Furthermore, if report cards use selected ICD-9-CM codes to indicate complications, the opposite of creep will occur: Hospitals and physicians will stop assigning these codes. In both situations, data quality suffers, as does the validity of the resultant quality measures. Implementation of regulatory controls (for example, penalties for certain coding practices) would not be feasible and would be expensive to enforce. Therefore, as administrative data are used to compare provider quality, more questions will arise about their validity for this purpose.

Administrative Data in the Future

Given the mixed appraisal of today's administrative data, what are the prospects for the future? I start with a safe prediction: Because of rapidly evolving information technologies, the definition, content, and scope of administrative data will change dramatically over the next several years. Large databases will contain extensive clinical information generated not only by clinicians and electronic reporting systems (such as those in laboratories and pharmacies) but also by patients. Instead of trying to fix existing administrative data for quality measurement, clinicians and researchers should develop a joint agenda to ensure the value of future databases for this purpose.

Boundaries between traditional administrative data and clinical information are blurring. Already,

information generated during patient care is transmitted automatically into electronic data repositories within some institutions. Admittedly, in most places, only "islands of automation" exist: digital information on health remains cloistered in departments or institutions (51). In the future, extensive electronic clinical data will be accumulated as a routine by-product of patient care, opening exciting possibilities for quality assessment. This could solve some data quality and definitional problems. For example, by knowing actual hematocrits, we could decide ourselves whether anemia is present; codes would be necessary only to specify type (such as sickle cell anemia or iron deficiency). By knowing hematocrits over time, we could infer the rapidity of blood loss.

Numerous efforts worldwide are shaping the future. Several U.S. and international bodies are establishing messaging standards that will permit extensive data exchanges from such sources as bedside computers, physiologic monitors, and billing systems throughout the health care delivery system. Determining the content of information to be exchanged, especially data supporting quality measurement, is a challenge to clinicians and researchers. Although the future seems bright, notes of caution are appearing. We need to answer four questions to move toward nonintrusive, electronically based systems for quality measurement.

What Should Data Sets Contain?

Arriving at a consensus about data content and standards for data transmission is under way in committees that represent the various stakeholders (for example, technology vendors, governments, medical representatives, and unions). These committees typically meet over several years, seeking external review and comment before submitting standards to a vote. Standards committees are accredited by such organizations as the American Society for Testing and Materials and the American National Standards Institute, but they are not always coordinated in their approaches (51).

In addition, fundamental disputes occasionally arise. For example, the American Medical Association (AMA) chairs the National Uniform Claim Committee, which was established in 1995 to develop a uniform set of data elements for electronic transmission of claims. Some committee members want claim forms to include data elements unavailable elsewhere, such as patients' self-reported functional status, specifically to support performance measurement. An AMA representative disagreed, observing, "Just because we're no longer restricted to an 8½-by-11-inch sheet of paper doesn't mean the sky's the limit" (52). Another AMA official concurred: "We don't want to burden physicians with

extra work or clutter up the claims form with superfluous information" (52).

These arguments crystallize the inevitable tension between data users and producers over the content of data sets—at least in the near future. One impediment is that the paradigm for data production has not yet shifted; many sectors still envision data production being done for specific, discrete tasks, such as billing. In the future, data will be generated (perhaps invisibly to both physician and patient) through each encounter with the health care delivery system. In addition, patients will enter self reports of functional status, satisfaction ratings, and other information directly into electronic systems.

How Should Data Be Coded?

Despite numerous reservations about ICD-9-CM, standards are essential for capturing diagnoses, procedures, and other nonnumeric clinical information. Otherwise, it is impossible to aggregate and compare data or to search databases for case findings or other information. The ICD-9-CM is one coding scheme, but others could also be used (for example, the new ICD-10, the *International Classification of Primary Care*, and the *International Classification of Impairments, Disabilities, and Handicaps*). Efforts are under way to define medical terms consistently (53), such as the Systematized Nomenclature of Medicine (SNOMED) project. The National Library of Medicine is combining SNOMED and more than 20 other standardized medical vocabularies into the Unified Medical Language System (UMLS). The UMLS Metathesaurus aims to link terms with common meanings, although even this massive undertaking is unlikely to satisfy all needs (54). Clinicians must participate in these efforts to ensure that ultimate coding schemes are clinically meaningful.

Several provisions of the Health Insurance Portability and Accountability Act of 1996 (Public Law 104-191), signed by President Clinton on 21 August 1996, give the federal government a leading role in determining standards for data transmission, mechanisms for protecting privacy, and coding methods. The law requires that the National Committee on Vital and Health Statistics "study the issues related to the adoption of uniform data standards for patient medical record information and the electronic exchange of such information" and provide "recommendations and legislative proposals for such standards and electronic exchange."

Who Will Produce Population-based Data?

Evaluating the quality of care across populations is of increasing interest, given constrained resources and shifting insurance plan enrollment. Rapidly developing telecommunications and networking technologies facilitate linkages of diverse, community-

wide data systems. Community health information networks and community health management information systems are designed to link public health officials, private medical practices, payers and insurers, and health care facilities within regions (51). Both share many features but have slightly different emphases. Community health information networks intend initially to facilitate connections and data transportation across organizations within regions, whereas community health management information systems typically concentrate on creating central data repositories for evaluating provider and health plan performance. Nontheless, both strive toward "seamless exchange of clinical or administrative information among health care providers, payers, and other authorized users" (51).

Approximately 100 community networks and information systems are being developed nationwide. Technology is usually not the limiting factor. Instead, progress is typically slowed by concerns about confidentiality and privacy, organization and owner control, and network design and data management. The most ambitious community health management information system to date, in Iowa, was delayed because of "overzealousness, poor planning . . . and conflicting agendas" (55). In addition, some are concerned that as managed care organizations increasingly become integrated health care delivery systems, the attraction of community networks and information systems will diminish (51). Not only will such integrated systems "wire" their extensive provider networks, competing organizations will refuse to share information. Others contend that large health care purchasers will force the public release of similar information across plans.

How Will Changes in Health Care Organization and Payment Affect Data?

Capitated payment plans, such as health maintenance organizations, are expanding rapidly in some health care markets. Capitated providers typically do not submit bills for individual services. For example, part B claims are not produced for Medicare beneficiaries who are enrolled in health maintenance organizations and part A hospital claims are submitted only for beneficiaries who are enrolled in cost-based health maintenance organizations (3). As more Medicare beneficiaries are encouraged to join health maintenance organizations, HCFA is currently undecided about what reporting standards to require. For example, should health maintenance organizations document services provided to Medicare beneficiaries during their year of enrollment? Although such data seem to be central to HCFA's fiduciary responsibility to ensure quality of care for Medicare beneficiaries, handling claims (even electronically) is costly.

Some health maintenance organizations argue forcefully against requiring reports about individual services, claiming that such requirements would increase administrative expenses and thus defeat a major purpose of capitation. Others recognize that some information about services is essential, if only for internal business purposes (such as planning, monitoring utilization, and profiling physicians). If health maintenance organizations face no external reporting requirements, databases developed for internal use are likely to diverge in content and scope, impeding comparisons across plans. In addition, health maintenance organizations may view internal data as proprietary and prohibit outside use or comparisons with other plans. Regardless, current trends in the financing and organization of health care raise questions about the availability of comparable administrative data from a burgeoning sector of the marketplace.

Conclusions

Current administrative data sets offer substantial practical advantages for widespread quality assessment. Limited clinical content and questions about data quality, however, compromise their utility for this purpose. Public report cards on provider performance derived from administrative data should highlight caveats that warn against drawing inappropriate conclusions from a single source. On the other hand, administrative data are useful as a screening tool for identifying quality problems and targeting areas that might require in-depth investigation. Administrative data in the future will rely extensively on electronic clinical databases, generating exciting opportunities for widespread quality assessment. Clinicians and researchers should actively participate in designing future administrative databases to ensure that they are clinically meaningful and useful for quality measurement.

References

1. **Wennberg JE, Gittelsohn A.** Small area variations in health care delivery. Science. 1973;182:1102-8.
2. **Center for Evaluative Clinical Studies, Dartmouth Medical School.** The Dartmouth Atlas of Health Care. Chicago: American Hosp Assoc; 1996.
3. **Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Mulbaier LH, et al.** Using Medicare claims for outcomes research. Med Care. 1994;32:JS38-51.
4. **Lave JR, Pashos CL, Anderson GF, Brailer D, Bubolz T, Conrad D, et al.** Costing medical care: using Medicare administrative data. Med Care. 1994; 32:JS77-89.

5. **Office of Technology Assessment, U.S. Congress.** Identifying Health Technologies That Work: Searching for Evidence. OTA-H-608. Washington, DC: US Gov Pr Off; 1994.

6. **Romano PS, Zach A, Luft HS, Rainwater J, Remy LL, Campa D.** The California Hospital Outcomes Project: using administrative data to compare hospital performance. Jt Comm J Qual Improv. 1995;21:668-82.

7. **Brook RH, Kamberg CJ, McGlynn EA.** Health system reform and quality. JAMA. 1996;276:476-80.

8. **Brook RH, McGlynn EA, Cleary PD.** Quality of health care. Part 2: measuring quality of care. N Engl J Med. 1996;335:966-70.

9. **Donabedian A.** Explorations in Quality Assessment and Monitoring. v I. The Definition of Quality and Approaches to Its Assessment. Chicago: Health Administration Pr; 1980.

10. **Lohr KN.** Outcome measurement: concepts and questions. Inquiry. 1988;25: 37-50.

11. **Iezzoni LI, ed.** Risk Adjustment for Measuring Healthcare Outcomes. 2d ed. Chicago: Health Administration Pr; 1997.

12. **Donaldson MS, Lohr KN.** Health Data in the Information Age. Use, Disclosure, and Privacy. Washington, DC: National Academy Pr; 1994.

13. **Health Care Financing Administration, Bureau of Data Management and Strategy.** Data Users Reference Guide. Baltimore: US Department of Health and Human Services; 1995.

14. **Health Care Financing Administration, Bureau of Data Management and Strategy.** Overview of Health Care Financing Administration Data. Resource Guide. Baltimore: US Department of Health and Human Services; 1996.

15. **Meistrell M, Schlehuber C.** Adopting a corporate perspective on databases. Improving support for research and decision making. Med Care. 1996;34: MS91-102.

16. **Epstein MH.** Guest alliance: uses of state-level hospital discharge databases. J Am Health Inform Manage Assoc. 1992;63:32-9.

17. **Ku L, Ellwood MR, Klemm J.** Deciphering Medicaid data: issues and needs. Health Care Financ Rev. 1990:35-45.

18. **Bright RA, Avorn J, Everitt DE.** Medicaid data as a resource for epidemiologic studies: strengths and limitations. J Clin Epidemiol. 1989;42:937-45.

19. **Garnick DW, Hendricks AM, Cornstock CB, Pryor DB.** A guide to using administrative data for medical effectiveness research. J Outcomes Manage. 1996;3:18-23.

20. **Gonnella JS, Hornbrook MC, Louis DZ.** Staging of disease. A case-mix measurement. JAMA. 1984;251:637-44.

21. **Young WW, Kohler S, Kowalski J.** PMC Patient Severity Scale: derivation and validation. Health Serv Res. 1994;29:367-90.

22. **Goldfield N, Boland P, eds.** Physician Profiling and Risk Adjustment. Gaithersburg, MD: Aspen; 1996.

23. **Kong D.** High hospital death rates. The Boston Globe. 3 October 1994:1,6,7.

24. **McMahon LF Jr, Smits HL.** Can Medicare prospective payment survive the ICD-9-CM disease classification system? Ann Intern Med. 1986;104:562-6.

25. **Payne TH, Murphy GR, Salazar AA.** How well does ICD-9 represent phrases used in the medical record problem list? Proc Annu Symp Comput Appl Med Care. 1993:654-7.

26. **Simborg DW.** DRG creep: a new hospital-acquired disease. N Engl J Med. 1981;304:1602-4.

27. **Waterstraat FL, Barlow J, Newman F.** Diagnostic coding quality and its impact on healthcare reimbursement: research perspectives. J Am Med Rec Assoc. 1990;61:52-9.

28. **Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP.** Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. N Engl J Med. 1988;318:352-5.

29. **Hsia DC, Ahern CA, Ritchie BP, Moscoe LM, Krushat WM.** Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. JAMA. 1992;268:896-9.

30. Feds hear medicine's plea, drop one Medicare form. Am Med News. 24 July 1995;38:26.

31. **Johnsson J.** Crackdown on coding ahead. Am Med News. 24 June 1996;39:28.

32. **Fowles JB, Lawthers AG, Weiner JP, Garnick DW, Petrie DS, Palmer RH.** Agreement between physicians' office records and Medicare part B claims data. Health Care Financ Rev. 1995;16:189-99.

33. **Wilson P, Smoley SR, Werdegar D.** Second Report of the California Hospital Outcomes Project. Acute Myocardial Infarction. Sacramento, CA: Office of Statewide Health Planning and Development; 1996.

34. **Jencks SF, Williams DK, Kay TL.** Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. JAMA. 1988; 260:2240-6.

35. **Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T.** Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? JAMA. 1992;267:2197-203.

36. **Romano PS, Mark DH.** Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care. 1994;32:81-90.

37. **Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD.** Using severity-adjusted stroke mortality rates to judge hospitals: how severity is measured may affect perceptions of hospital performance. Int J Qual Health Care. 1995;7:81-94.

38. **Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD.** Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. Ann Intern Med. 1995;123:763-70.

39. **Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD.** Severity measurement methods and judging hospital death rates for pneumonia. Med Care. 1996;34:11-28.

40. **Landon B, Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, et al.** Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. Inquiry. 1996;33:155-66.

41. **Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD.** Using severity measures to predict likelihood of death for pneumonia inpatients. J Gen Intern Med. 1996;11:23-31.

42. **Steen PM, Brewster AC, Bradbury RC, Estabrook E, Young JA.** Predicted probabilities of hospital death as a measure of admission severity of illness. Inquiry. 1993;30:128-41.

43. **Hannan EL, Racz MJ, Jollis JG, Peterson ED.** Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? Health Serv Res. 1997;31:659-78.

44. **Hughes JS, Iezzoni LI, Daley J, Greenberg L.** How severity measures rate hospitalized patients. J Gen Intern Med. 1996;11:303-11.

45. **Shapiro MF, Park RE, Keesey J, Brook RH.** The effect of alternative case-mix adjustments on mortality differences between municipal and voluntary hospitals in New York City. Health Serv Res. 1994;29:95-112.

46. **Roos LL, Walld R, Wajda A, Bond R, Hartford K.** Record linkage strategies, outpatient procedures, and administrative data. Med Care. 1996;34: 570-82.

47. **Iglehart JK.** The National Committee for Quality Assurance. N Engl J Med. 1996;335:995-9.

48. **Weiner JP, Parente ST, Garnick DW, Fowles J, Lawthers AG, Palmer RH.** Variation in office-based quality. A claims-based profile of care provided to Medicare patients with diabetes. JAMA. 1995;273:1503-7.

49. **Koska M.** Are severity data an effective consumer tool? Journal of the American Hospital Association. 1989;63:24.

50. **Hannan EL, Kilburn H Jr, Lindsey ML, Lewis R.** Clinical versus administrative data bases for CABG surgery. Does it matter? Med Care. 1992;30:892-907.

51. **Office of Technology Assessment, U.S. Congress.** Bringing Health Care Online. The Role of Information Technologies. OTA-ITC-624. Washington, DC: US Gov Pr Off; September 1995.

52. **Borzo G.** AMA wants role in evolving system for billing payers. Am Med News. 22 July 1996;39:7-8.

53. **Cimino JJ.** Review paper: coding systems in health care. Methods Inf Med. 1996;35:273-84.

54. **Payne TH, Martin DR.** How useful is the UMLS metathesaurus in developing a controlled vocabulary for an automated problem list? Proc Annu Symp Comput Appl Med Care. 1993;705-9.

55. **Borzo G.** Iowa delays, scales back electronic billing plan. Am Med News. 29 July 1996;39:4-5.