Agreement Between Physicians' Office Records and Medicare Part B Claims Data

Jinnet B. Fowles, Ph.D., Ann G. Lawthers, Sc.D., Jonathan P. Weiner, Dr.P.H., Deborah W. Garnick, Sc.D., Doris S. Petrie, and R. Heather Palmer, M.B., B.Ch., S.M.

This article tests agreement between demographic, diagnostic, and procedural information from primary-care physicians' office records and Medicare Part B claims for Maryland Medicare beneficiaries. The extent of agreement depended on the category of information being compared. Demographics matched poorly, probably due to incomplete record samples. Important diagnoses were often missing from the medical record. When claims indicated presence of disease, the patient was likely to have the disease, but claims did not capture all people who have the disease. Additonally, many laboratory tests and procedures were missing from the primary-care record. The appropriate use of either of these data sources depends on the specific research question that is being asked.

INTRODUCTION

Both payers and providers agree that measures of quality are the linchpin of successful health care reform. In the 1990s, health care payers are eager to provide their customers with access to the most cost-effective health care systems. To reach this goal, they must be able to accurately compare costs and performance of competing systems. Providers need to be able to measure their own performance as part of the process of improving it. Both groups need accurate data to achieve their objectives.

Two data sources frequently used to measure quality of care are medical records and administrative billing sources (i.e., claims). The development of diagnosis-related groups (DRGs) in the 1980s triggered extensive investigation of the accuracy of hospital medical records (Roos, Sharp, and Wajda, 1989; Hsia et al., 1988; Lloyd and Rissing, 1985; Feigl et al., 1988; Iezzoni et al., 1988) and use of hospital billing and discharge data (Wennberg et al., 1987; Roos et al., 1985, 1990; Luft and Hunt, 1986; Anderson et al., 1990; Iezzoni et al., 1992; Maronde et al., 1989; Romano, 1993). Hospital medical records have been the locus of extensive quality assurance activities and billing verification. In fact, hospital records have become the gold standard against which other sources, such as claims data and patient survey information, are compared (Roos et al., 1982; Leatherman et al., 1991).

Less is known about the quality of ambulatory medical records and claims (Palmer, 1988; Horner et al., 1991; Lurie et al., 1992). In contrast to hospital medical records, physicians' office medical records have escaped such careful scrutiny (Buchsbaum, Boling, and Groh, 1987; Moran et al., 1988; Romm and Putnam, 1981). Similarly, with few exceptions (Quam et al., 1993; Graft et al., 1992), billing information from ambulatory claims has also been studied less frequently. Because the volume and costs of

The research presented in this article was supported by the Health Care Financing Administration (HCFA) under Contract Number 500-89-0624. Jinnet B. Fowles is with the Park Nicollet Medical Foundation. Ann G. Lawthers and R. Heather Palmer are with the Harvard School of Public Health. Jonathan P. Weiner is with the Johns Hopkins School of Public Health. Deborah W. Garnick is with the Heller Graduate School, Brandeis University. Doris S. Petrie was with the Park Nicollet Medical Foundation when this research was conducted; she has since retired. The opinions expressed are those of the authors and do not necessarily reflect those of HCFA, the Park Nicollet Medical Foundation, Harvard School of Public Health, Johns Hopkins School of Public Health, or Brandeis University.

ambulatory medical care are increasing, it is critical to understand the strengths and weaknesses of ambulatory-care data sources. This article compares information from physicians' office medical records with information from claims data for a population of Medicare beneficiaries.

METHODS

The data for this study come from a larger project called "Developing and Evaluating Methods to Promote Ambulatory Care Quality" (DEMPAQ). That project, which has been described elsewhere (Lawthers et al., 1993, 1995; Garnick et al., 1994a, 1994b; Parente et al., in press), tested two methods for reviewing office-based care given to Medicare beneficiaries. DEMPAQ was funded by HCFA to evaluate techniques that could be used to implement a congressional mandate for office-based review. One method, using medical-record review, was based on activities that physicians typically perform in the course of an office visit, such as drug monitoring. The second method used ambulatory and hospital claims to profile practice patterns for chronic diseases, prevention, and general office practice. This study compares Medicare Part B claims data from the office setting with office medical-record data from a sample of primary-care physicians and their patients.

MEDICAL-RECORD DATA SOURCES

The study sample was identified from Maryland Medicare beneficiaries who had received billed medical services from a primary-care physician between July 1, 1989, and June 30, 1990. Each beneficiary was assigned to an internist, family practitioner, or general practitioner who had provided the majority of their visits. A random sample (n = 300) of all Maryland physicians who had 25 or more Medicare beneficiaries assigned to them (n = 1,329) were invited by letter to submit copies of a 2-year portion (July 1, 1989–June 30, 1991) of their medical records for a random subset of 20 of their patients. The physician response rate was 30.3 percent (n = 91), representing the records of 1,998 patients. Non-respondents were called at least once by the Delmarva Foundation for Medical Care, Inc., the Maryland peer review organization (PRO). Because HCFA was committed to testing a voluntary program and because recruitment time for the project was limited, no further followup of non-respondents was done.

To avoid bias, reviewers who abstracted the records did not know which codes were on the claims. We trained eight medical-record review nurses from the Delmarva PRO to use a computer-based medical-record abstraction system to search for a specified number of items, including demographic information, diagnoses, and laboratory tests and procedures. These items had been selected on the basis of their inclusion in the claims-based practice profiles. Synonym dictionaries were created and imbedded in the computer-based medical-record review format. It is important to note that subjective clinical judgments were not permitted. For example, reviewers were not allowed to assume that elevated blood pressure readings constituted a diagnosis of hypertension if the physician had not specifically indicated that diagnosis. After the reviewers' training period, there was an ongoing 5-percent re-review, with one randomly selected medical record from each physician assigned to pairs of reviewers. Interrater reliability exceeded 95 percent.

CLAIMS AND ELIGIBILITY DATA SOURCES

A parallel set of all claims, including both institutional Parts A and B and non-institu-

tional Part B claims for these 1,998 patients, was assembled from the HCFA National Claims History file. The raw claims were cleaned and organized, removing repetitive adjudication data and constructing a file that contained a single line for each unique service. Claims data were assembled in January 1992, allowing ample time for all claims to be filed that were related to the review period for this analysis (July 1, 1990-June 30, 1991). While claims-based analyses for the original DEMPAQ study used both Part A and Part B claims, this reliability study used only Part B claims. It was in the Part B files that we found claims submitted by the primary-care provider.

In addition to the claims, we also abstracted information from the HCFA Health Insurance Skeleton Eligibility Write-Off (HISKEW) files. This information included the Medicare beneficiary identifier, date of birth, gender, and ZIP Code.

LINKING CLAIMS AND RECORDS

Because the claims analysis for the larger DEMPAQ project was based on the second 12-month period, this study of agreement was restricted to the same 12 months (July 1, 1990–June 30, 1991). To compare claims data with medical record data, we created a translation table, specifying the codes from the *Current Procedural Terminology, 4th Edition* for procedures and the *International Classification of Diseases, 9th Revision, Clinical Modification* for diagnoses to be considered as matches in the claims files for the items searched in the medical record. These coding definitions were designed to be inclusive. For example, in the medical record, we searched for evidence of a blood urea nitrogen test. Since such tests often are ordered as part of a test panel, we accepted panel tests in the claims portion as reasonable indicators that the specific test was performed. This rule is biased in favor of over-identifying matches between claims and medical-record information for laboratory tests and procedures.

For matching claims-based diagnoses with those in the primary-care physician's medical record, we considered only those diagnoses that occurred on bills submitted by the primary-care physician (Part B claims with the physician's Unique Provider Identification Number). In contrast, we matched laboratory tests and procedures irrespective of the billing entity, since most of these services are provided by a laboratory, even though the tests may have been ordered by the primary-care physician.

Analysis

To compare events (diagnoses, laboratory tests, and procedures) between the medical record and the claim, we calculated both percent agreement and the kappa *(K)* statistic (Rosner, 1986). The percent agreement is inflated because of the large number of cases in which the event does not occur in either the medical record or the claims. The *K* coefficient of agreement is the ratio of the proportion of times that the sources agree (corrected for chance agreement) to the maximum proportion of times that the sources could agree (corrected for chance agreement). Its value can range from -1.0 to 1.0. While there is no predetermined level of acceptability, some suggest that a *K* of less than 0.40 indicates a poor level of agreement (Fleiss, 1981).

RESULTS

Of the 1,998 medical records we received from primary-care physicians, we were able to identify activity on the claims files for 1,927. Presumably, the 71 other patients had received services in the year

used to sample cases but not in the subsequent study year. Analyses for demographics and tests/procedures are based on those 1,927 medical records. Of the 1,927 medical records, only 1,596 had indications of any visit to the primary-care physician in the medical record during the 12-month study period; therefore, the analysis of diagnoses is based on those 1,596 cases.

Demographics

As shown in Table 1, the data from the HISKEW files and the medical record agree well for name and gender only. Name matched in 99.4 percent of cases (when misspellings are considered as matches), and gender matched in 97.7 percent of cases. For all other demographic data, there is a striking lack of agreement. The patient ZIP Code matched in only 40.3 percent of cases, and date of birth matched in 58.5 percent of cases. Determination of whether the patient was alive at the end of the review period or had died (death status) agreed in 34.3 percent of cases, even given that reviewers were allowed to infer that medical-record activity after the end of the study period constituted evidence of survival through the study period.

Diagnosis

We searched the medical record for the explicit mention of any of 27 different diagnoses anytime during the study period. Only 1,596 medical records listed any physician visits at all during the study period. Matches were not made for specific dates, but rather we defined as a match any code of the diagnosis in the claims submitted by the primary-care physician during the 12-month study period, compared with any verbatim mention in the medical record in the same timeframe. The percent agreement ranged from 74.9 percent for hypertension to 99.7 percent for ketoacidosis (Table 2). However, the percent agreement is inflated because many patients did not have the disease, thus increasing the match of "not in records nor in claims." As shown in Figure 1, the range for the *K* statistic is from 0.0 to 0.72. *K* met or exceeded 0.40 for only six of the diagnoses. The best agreement, as measured by *K*, was for the most serious major chronic diseases (congestive heart failure, chronic obstructive pulmonary disease [COPD], diabetes, hypertension, rheumatoid arthritis, and transient ischemic attack).

Table 1

Agreement Between Demographic Characteristics in Physicians' Medical Records and Medicare HISKEW Files

Characteristic	*n*	Percent
Name		
Match	1,854	96.2
No Match	73	3.8
Misspelling	62	
Missing in HISKEW	8	
Complete Mismatch	3	
Gender		
Match	1,883	97.7
No Match	44	2.3
2 in HISKEW, 1 in Medical Record	19	—
1 in HISKEW, 2 in Medical Record	12	—
Missing in Medical Record	13	—
ZIP Code		
Match	776	40.3
No Match	1,151	59.7
In Medical Record, No Match	87	4.5
Missing in Medical Record	1,064	55.2
Date of Birth		
Match	1,127	58.5
No Match	800	41.5
In Medical Record, No Match	210	10.9
Missing in Medical Record	590	30.6
Death		
Match	660	34.3
No Match	1,267	65.7
Missing Review Data	2	0.1
Unknown in Medical Record	1,260	65.4
In Medical Record, Not in HISKEW	3	0.2
In HISKEW, Not in Medical Record	2	0.2

NOTES: HISKEW is Health Insurance Skeleton Eligibility Write-off files. *n* = 1,927. Sample includes Maryland Medicare beneficiaries with billed services between July 1, 1990 and June 30, 1991.

SOURCE: Fowles J.B., Park Nicollet Medical Foundation, Lawthers, A.G., and Palmer, R.H., Harvard School of Public Health, Weiner, J.P., Johns Hopkins School of Public Health, Garnick, D.W., Brandeis University, and Petrie, D.S., 1995.

Table 2

Agreement Between Diagnosis in Physicians' Medical Records[1] and Medicare Part B Claims[2]

Diagnosis	Agreement: In Record and Claim	Agreement: Not in Records Nor in Claims	Disagreement: In Claim Not in Records	Disagreement: In Record Not in Claims	Percent Agreement
Gangrene	0	1,596	0	0	100.0
Ketoacidosis	0	1,591	4	1	99.7
Rheumatoid Arthritis	4	1,582	5	5	99.4
Acute Myocardial Infarction	3	1,582	8	3	99.3
Hypokalemia	0	1,585	4	7	99.3
Diabetic Retinopathy	0	1,584	2	10	99.2
Transient Ischemic Attack	5	1,576	4	11	99.1
Cardiomegaly	1	1,581	0	14	99.1
Renal Failure	1	1,577	7	11	98.9
Pneumonia	6	1,569	14	7	98.7
Influenza	4	1,566	16	10	98.4
Stroke	5	1,562	5	24	98.2
Peptic or Gastric Ulcer	6	1,560	9	21	98.1
Neuropathy	3	1,533	4	56	96.2
Congestive Heart Failure	58	1,468	29	41	95.6
Angina	23	1,500	18	55	95.4
Chronic Obstructive Pulmonary Disease	74	1,416	37	69	93.4
Diabetes Mellitus	175	1,313	54	54	93.2
Chest Pain	21	1,467	32	76	93.2
Gastrointestinal Upset	16	1,464	11	105	92.7
Bronchitis, Acute	23	1,447	25	101	92.1
Hyperlipidemia	32	1,411	37	116	90.4
Upper Respiratory Infection	52	1,365	19	160	88.8
Osteoarthritis	59	1,343	71	123	87.8
Arthritis, Unspecified	46	1,341	31	178	86.9
Ischemic Heart Disease	55	1,331	100	110	86.8
Hypertension	361	835	203	197	74.9

[1] Diagnosis recorded at a visit at least once.
[2] Face-to-face visits billed by that physician.

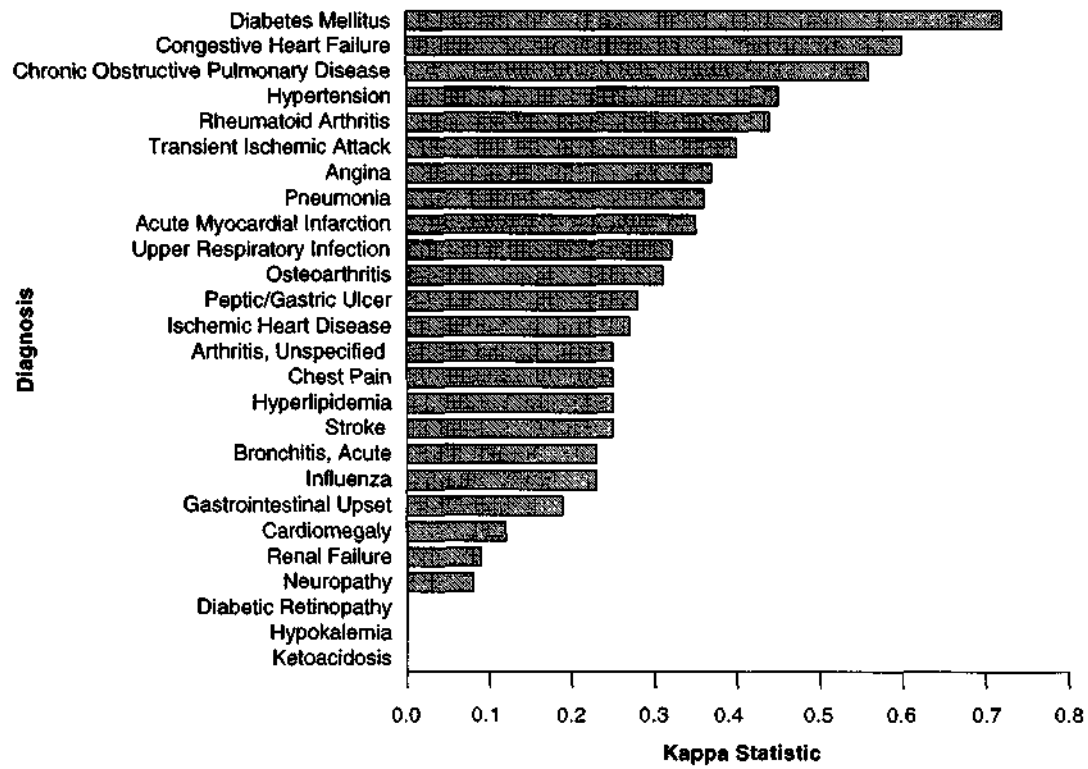
NOTES: n = 1,596. Sample includes Maryland Medicare beneficiaries with billed services between July 1, 1990 and June 30, 1991.

SOURCE: Fowles J.B., Park Nicollet Medical Foundation, Lawthers, A.G., and Palmer, R.H., Harvard School of Public Health, Weiner, J.P., Johns Hopkins School of Public Health, Garnick, D.W., Brandeis University, and Petrie, D.S., 1995.

Generally, diagnoses appeared more often in the record than in the claims. Notable exceptions to this pattern included potentially severe acute conditions that appeared more often in the claim than in the medical record, such as acute myocardial infarction, diabetes, hypertension, ischemic heart disease, and pneumonia. These mismatches surprised us, since we had been expecting severe diagnoses to always be noted in the primary-care physician's record. Therefore, we conducted a re-review of a small sample of these records. Two reviewers (not among the original 8) independently reviewed 21 medical records. The records were selected because claims indicated that the beneficiary had a disease, while the original medical-record review indicated that the beneficiary did not. Six diseases of special interest (diabetes, congestive heart failure, COPD, hypertension, ischemic heart disease, and osteoarthritis) were included. For re-review, the reviewers knew the specific diagnosis that the claims had indicated, and the reviewers were allowed to use supporting clinical evidence to answer the question, "Does this physician think this patient has this disease?" Acceptable evidence included a mention of the specific diagnosis before the 12-month study period or other clinical indicators, such as prescribing micronase and following blood sugars as substantiation of diabetes. When the two reviewers did not agree (as happened in two cases), the case was

Figure 1

Kappa Statistic for Agreement Between Diagnosis in Physicians' Medical Records and Medicare Part B Claims



NOTES: n = 1,596. Sample includes Maryland Medicare beneficiaries with billed services between July 1, 1990 and June 30, 1991.

SOURCE: Fowles J.B., Park Nicollet Medical Foundation, Lawthers, A.G., and Palmer, R.H., Harvard School of Public Health, Weiner, J.P., Johns Hopkins School of Public Health, Garnick, D.W., Brandeis University, and Petrie, D.S., 1995.

reviewed by a family practitioner whose judgment was accepted.

In almost one-half of the cases, the diagnosis had been noted in the previous 12 months. In three additional cases, there was clinical evidence indicating that the physician was treating the patient for that disease. Data-abstraction errors were relatively uncommon and went in both directions; that is, in two cases the initial reviewer had missed the diagnosis, and in two cases the claims assignment appeared unwarranted. In fact, 81 percent of all cases with no verbatim mention of the diagnosis in the medical record were verified as having the condition noted on the claim, when this broader definition of a match was used. In 9.5 percent of the cases, the condition listed in the claims was not supported with any evidence in the medical record; the final 9.5 percent of cases were indeterminate because the records were illegible.

Laboratory Tests and Procedures

In contrast with our uncertainty about the validity of diagnoses in claims or medical records, we can regard the claims as a reasonable gold standard for whether or not the patient received any laboratory test or procedure. The strength of this conviction rests on the fact that physicians are paid only for procedures for which they submit bills. Accurate diagnosis is not required for payment.

Table 3

Agreement Between Lab Tests and Procedures in Physicians' Medical Records and Medicare Part B Claims

Lab Test or Procedure	*K*	Agreement: In Record and Claim	Agreement: Not in Records Nor in Claim	Disagreement: In Claim Not in Records	Disagreement: In Record Not in Claims	Percent Agreement
Hemoglobin A1C	0.73	54	1,835	24	14	98.0
Serum Creatinine	0.56	608	902	266	151	78.4
Blood Urea Nitrogen	0.55	616	888	268	155	78.0
Serum Calcium	0.55	572	936	293	126	78.3
Potassium	0.55	628	871	273	155	77.8
Sodium	0.55	600	903	272	152	78.0
Mammogram	0.53	126	1,614	171	16	90.3
Serum Glucose	0.52	674	784	317	152	75.7
Continuous Oxygen Therapy	0.50	9	1,900	17	1	99.1
Influenza Shot	0.36	113	1,528	87	199	85.2
Spine X-Ray	0.36	41	1,753	124	9	93.1
Electrocardiogram	0.34	325	988	584	30	68.1
Exercise Tolerance Test	0.33	21	1,826	74	6	95.8
Pulmonary Function Tests	0.31	17	1,837	70	3	96.2
Nuclear Medicine Cardiac Stress Test	0.30	9	1,877	37	4	97.9
Pap Smear	0.29	29	1,773	90	35	93.5
Flexible Sigmoidoscopy	0.28	11	1,861	49	6	97.1
Colonoscopy	0.27	15	1,835	76	1	96.0
Chest X-Ray	0.25	171	1,155	589	12	68.8
Hip X-Ray	0.24	13	1,836	77	1	96.0
Knee X-Ray	0.24	14	1,828	82	3	95.6
Finger X-Ray	0.22	1	1,919	7	0	99.6
Foot X-Ray	0.22	10	1,851	65	1	96.6
Echocardiogram	0.16	18	1,741	167	1	91.3
Blood Gases	0.13	4	1,870	50	3	97.2
Hand X-Ray	0.06	1	1,894	31	1	98.3
Debridement	0.04	1	1,880	41	5	97.6
Echo Stress Test	0.00	0	1,918	6	3	99.5
Ejection Fraction	0.00	0	1,924	0	3	99.8
Toe X-Ray	0.00	0	1,923	4	0	99.8
Nuclear Medicine Test	-0.02	0	1,857	30	40	96.4

NOTES: n = 1,927. Sample includes Maryland Medicare beneficiaries with billed services between July 1, 1990 and June 30, 1991.

SOURCE: Fowles J.B., Park Nicollet Medical Foundation, Lawthers, A.G., and Palmer, R.H., Harvard School of Public Health, Weiner, J.P., Johns Hopkins School of Public Health, Garnick, D.W., Brandeis University, and Petrie, D.S., 1995.

As shown in Table 3, percent agreement ranged from 68.1 percent to 99.8 percent, and *K* ranged from -0.02 to 0.73. In general, the correspondence between claims and medical records is worse for tests and procedures than for diagnoses. Not unexpectedly, tests and procedures appeared more frequently in the claims than in the medical record. However, it is more difficult to interpret the "in claim, not in records" information because of billing patterns. Claims do not currently carry information about who ordered the test or procedure, only who billed for the test. Therefore, tests could have been ordered by some other physician and not found in the primary-care physicians' office medical record. Additional explanations may include that the test was not filed in the medical record, that the report was not copied for review, or even that there was fraud and the test was billed but not done. In any case, these results suggest that much testing is being conducted of which the primary-care physician is probably not aware.

The other type of mismatch — in which we have evidence of the test in the medical record but not in the claims — is more problematic. We commonly believe that a physician would submit a claim for any billable procedure because there is a financial incentive to do so. Several plausible expla-

nations for why a bill would not exist for a procedure or laboratory test that was done include:

- The tests could have been done in the hospital and grouped as part of a revenue code.
- A bill would not be sent to Medicare for tests or procedures that Medicare does not cover (e.g., at the time of the study, screening mammograms or influenza immunizations).
- A physician or laboratory forgot to submit a bill.
- There were errors in entering the bill into the Medicare system.

DISCUSSION

From the results of both our principal and followup analyses of claims-based diagnoses, we conclude that when the claim indicates that a patient has a disease, the person probably has the disease. Our second, unblinded, and more subjective review allowed us to identify a number of diagnoses in the medical record that the initial, blinded, objective review had missed. However, the claims do not identify all people with the disease. We conclude that ambulatory claims are less useful in documenting the prevalence or incidence of diseases because they understate the disease rates. There were many instances of disease being documented in the medical record but never appearing in the Part B claims. Therefore, the usefulness of claims depends on the nature of the question being asked.

For many quality assurance purposes, claims can be an important resource, since we are often more concerned with rates of certain procedures being done for patients with given conditions (e.g., the number of people with diabetes who had a hemogloblin A1c), rather than the total number of people who had the condition. From quality of care guidelines, we can identify at least a subset of the populations of relevance (the denominator) and many of the processes of care (the numerator). Both the numerator and denominator can be identified using claims information. When measuring quality improvement, we are often looking at rates of performance, and we prefer to give providers the benefit of the doubt. We do not want to inflate the denominator with inappropriate cases, because the result would be a lower performance rate than is warranted. An example of how powerful this type of quality-of-care analysis can be is demonstrated in a recent article using data from the original DEMPAQ study (Weiner et al., 1995).

The findings of this research support the conduct of studies in the quality of care. Such studies must increasingly focus on the care delivered in primary-care settings. Primary-care sites are an essential component of the health care delivery system because of the frequency of interactions that occur in that setting and the seminal nature of such interactions. Adequate primary care can avert more costly hospitalization through the delivery of preventive services and the management of chronic diseases. Furthermore, management of post-hospital care increasingly occurs in the primary-care setting. With the emergence of managed care and the reduction in hospitalization rates and length of stay, primary-care office settings move to the forefront of quality-of-care research.

Other quality-of-care initiatives also depend on interpretable primary-care data. For example, there is a surging volume of guidelines aimed at improving the quality of care and reducing practice variation. To be effective, these guidelines need adequate measures of performance for giving feedback to providers. Feedback information depends on having understandable

and accessible measures of the processes and outcomes of care. Providers need to understand the strengths and limits of these measures before they can accept practice profile results and work to improve their care.

Caution needs to be exercised in interpreting and generalizing from these results for three reasons. First, these records come from primary-care physicians only. Specialists may have different documentation practices. Second, these analyses are based on records from volunteer physicians who may have better medical records than non-volunteers.[1] At a minimum, the medical records from volunteers were more likely to be legible, and the volunteer physicians felt comfortable providing their records for use in developing measures of the quality of health care. Third, the diagnosis-recording habits of physicians in solo or small group practices in Maryland may not be generalizable to all physicians in all practice settings. For example, in a similar study conducted in a large midwestern multi-specialty group practice, verbatim mention of a chronic disease was missing in only 1 of 200 cases (Health Research Center, 1994). Further research is required to determine the variation and associated factors in documentation patterns.

The medical record did not provide a good demographic profile, but this weakness may have been a function of the initial request made to participating physicians. We had requested transactional information for a 2-year period (visits, laboratory notes, X-ray reports, consultation notes, hospitalizations, etc.). Most of the lack of matching for demographic characteristics was due to missing information in the 12-month portion of the medical record that was reviewed. This finding suggests that the visit section of the medical record alone is an insufficient source for providing demographic information. The primary-care physician possibly maintains separate registration files that contain some of the missing elements. Alternatively, such information may have been recorded on a medical-record facesheet that was not submitted. Similarly, an explicit request for information subsequent to the close of the analysis period would have yielded information allowing us to conclude that the patient was still alive.

In contrast to hospital medical records, ambulatory records cannot now be considered a reasonable gold standard for diagnoses, testing, or procedures, because the records often lack complete documentation. For example, when a diagnosis is missing from the medical record but appears in the claims, despite no verbatim documentation of diagnosis, the disease often is being actively managed. Moreover, common, serious chronic diseases are frequently not explicitly named in these physicians' office records.

This documentation pattern is distinctly different from that in hospital records and requires a different review strategy. To assess whether the physician is treating the patient for a particular disease, the reviewers must be trained to use consistent clinical judgment to draw conclusions about diagnoses. Such a strategy requires a higher level of clinical expertise and training to achieve valid and reliable results. These knowledge and training requirements, in turn, make reviewing office records more expensive than usually is anticipated.

Clearly, Medicare claims can indicate much activity of which the primary-care physician may not be aware. The results of

[1] As part of the original DEMPAQ project, we conducted a separate comparison of the practice characteristics of volunteer and non-volunteer physicians. Using claims data to compare the two groups, we did not find significant differences between volunteer and non-volunteer physicians in terms of age, type of practice (solo versus group), geographic location (urban versus rural), quality of care, and hospital use. General practitioners were less likely to participate than family practitioners or internists (Delmarva Foundation for Medical Care, Inc., 1993).

this study suggest that the role of primary-care physician as the coordinator of care is not being fulfilled. One can ask whether or not the total numbers of procedures would be reduced if this coordination function were to be achieved. Further research is required to assess the cost-effectiveness of coordinated care, and could be done with data like those described in this article. Separate studies examining the medical records of specialists with claims or the medical records of all providers with claims could offer a more complete picture of the use of ambulatory-care claims.

Our finding points to potential new applications of claims for alerting physicians to other care that is being rendered. A common electronic medical record that is accessible to all providers would provide such information. Until claims and medical-record data merge in an electronic medical record, we must continue to use the available information sources to inform our judgments about the quality of care. These judgments will be wiser when tempered with knowledge about the strengths and limits of the data for specific applications.

ACKNOWLEDGMENTS

The authors thank Jean Edwards, Sandy Moentmann, Stephen Parente, Paul Chandler, and Thomas Richards for their substantial contributions to the testing process.

REFERENCES

Anderson, G., Steinberg, E.P., Whittle, J., et al.: Development of Clinical and Economic Prognoses From Medicare Claims Data. *Journal of the American Medical Association* 263(7):967-972, February 16, 1990.

Buchsbaum, D.G., Boling, P., and Groh, M.: Residents' Underdocumentation in Elderly Patients' Records of Prescriptions for Benzodiazepine. *Journal of Medical Education* 62:438-440, May 1987.

Delmarva Foundation for Medical Care, Inc.: *Developing and Evaluating Methods to Promote Ambulatory Care Quality (DEMPAQ):* Final report Prepared for the Health Care Financing Administration. Contract Number 500-89-0624. Easton, MD. 1993.

Feigl, P., Glaefke, G., Ford, L., et al.: Studying Patterns of Cancer Care: How Useful Is the Medical Record? *American Journal of Public Health* 78(5):526-533, May 1988.

Fleiss, J.L.: *Statistical Methods for Rates and Proportions, 2nd Edition.* New York. John Wiley & Sons, 1981.

Garnick, D.W., Fowles, J., Lawthers, A.G., et al.: Focus on Quality: Profiling Physicians' Practice Patterns. *Journal of Ambulatory Care Management* 17(3):44-75, 1994a.

Garnick, D.W., Lawthers, A.G., Moentmann, S. et al.: An Affordable System for Reviewing Medical Records From Physicians' Offices. *Joint Commission Journal on Quality Improvement* 20(12):679-694, 1994b.

Graft, D.F., Fowles, J., McCoy, C.E., and Lager, R.A.: Detection of Beta-Blocker Use in People With Asthma. *Annals of Allergy* 69:449-453, November 1992.

Health Research Center, Park Nicollet Medical Foundation: *Alternative Approaches to Risk Measurement: Final Report to Physician Payment Review Commission.* Grant Number 93-G07. Minneapolis, Minnesota 1994.

Horner, R.D., Paris, J.A., Purvis, J.R., and Lawler, F.H.: Accuracy of Patient Encounter and Billing Information in Ambulatory Care. *Journal of Family Practice* 33(6):593-598, 1991.

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(6):352-355, February 11, 1988.

Iezzoni, L.I., Burnside, S., Sickles, L., et al.: Coding of Acute Myocardial Infarction: Clinical and Policy Implications. *Annals of Internal Medicine* 109:745-751, November 1, 1988.

Iezzoni, L.I., Foley, S.M., Daley, 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* 267(16):2197-2203, April 22/29, 1992.

Lawthers, A.G., Palmer, R.H., Edwards, J.E., et al.: Developing and Evaluating Performance Measures for Ambulatory Care Quality: A Preliminary Report of the DEMPAQ Project. *Joint Commission Journal on Quality Improvement* 19(12):552-565, 1993.

Lawthers, A.G., Palmer, R.H., Garnick, D.W., et al.: Designing and Using Measures of Quality Based on Physician Office Records. *Journal of Ambulatory Care Management* 18(1):56-72, 1995.

Leatherman, S., Peterson, E., Heinen, L., and Quam, L.: Quality Screening and Management Using Claims Data in a Managed Care Setting. *Quality Review Bulletin* Pp. 349-359, November 1991.

Lloyd, S.S., and Rissing, J.P.: Physician and Coding Errors in Patient Records. *Journal of the American Medical Association* 254(10):1330-1336, September 13, 1985.

Luft, H.S., and Hunt, S.S.: Evaluating Individual Hospital Quality Through Outcome Statistics. *Journal of the American Medical Association* 255(20):2780-2784, May 23/30, 1986.

Lurie, N., Popkin, M., Dysken, M., et al.: Accuracy of Diagnoses of Schizophrenia in Medicaid Claims. *Hospital and Community Psychiatry* 43(1):69-71, January 1992.

Maronde, R.F., Chan, L.S., Larsen, F.J., et al.: Underutilization of Antihypertensive Drugs and Associated Hospitalization. *Medical Care* 27(12):1159-1166, December 1989.

Moran, M.T., Wiser, T.H., Nanda, J., and Gross, H.: Measuring Medical Residents' Chart-Documentation Practices. *Journal of Medical Education* 63:859-865, November 1988.

Palmer, R.H.: The Challenges and Prospects for Quality Assessment and Assurance in Ambulatory Care. *Inquiry* 25:119-131, Spring 1988.

Parente, S., Weiner, J.P., Garnick, D.W. et al.: Developing a Quality Improvement Database Using Health Insurance Data: A Guided Tour With Application to Medicare's National Claims History File. *American Journal of Medical Quality.* In press.

Quam, L., Ellis, L.B.M., Venus, P., et al.: Using Claims Data for Epidemiologic Research: The Concordance of Claims-Based Criteria With the Medical Record and Patient Survey for Identifying a Hypertensive Population. *Medical Care* 31(6):498-507, 1993.

Romano, P.S.: Can Administrative Data Be Used to Compare the Quality of Health Care? *Medical Care Review* 50(4):451-477, Winter 1993.

Romm, F.J., and Putnam, S.M.: The Validity of the Medical Record. *Medical Care* 19(3):310-315, March 1981.

Roos, L.L., Fischer, E.S., Sharp, S.M., et al.: Postsurgical Mortality in Manitoba and New England. *Journal of the American Medical Association* 263(18):2453-2458, May 9, 1990.

Roos, L.L., Cageorge, S.M., Austen, E., and Lohr, K.N.: Using Computers to Identify Complications After Surgery. *American Journal of Public Health* 75(11):1288-1295, November 1985.

Roos, L.L., Roos, N.P., Cageorge, S.M., and Nicol, J.P.: How Good Are the Data? Reliability of One Health Care Data Bank. *Medical Care* 20(3):266-276, March 1982.

Roos, L.L., Sharp, S.M., and Wajda, A.: Assessing Data Quality: A Computerized Approach. *Social Science and Medicine* 28(2):175-182, 1989.

Rosner, B.: *Fundamentals of Biostatistics, 2nd Edition.* Boston. Duxbury Press, 1986.

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 Association* 257(7):933-936, February 20, 1987.

Weiner, J., Parente, S., Garnick, D., et al.: Variation in Office-Based Quality: A Claims-Based Profile of Care to Medicare Patients With Diabetes. *Journal of the American Medical Association* (273) 19:1503-1508, May 17, 1995

Reprint Requests: Jinnet B. Fowles, Ph.D., Park Nicollet Medical Foundation, 3800 Park Nicollet Boulevard, Minneapolis, Minnesota 55416.