Accuracy of Patient Encounter and Billing Information in Ambulatory Care

Ronnie D. Horner, PhD, Janice A. Paris, MSPH, John R. Purvis, MD, and Frank H. Lawler, MD

Greenville, North Carolina

Background. This study examined the degree of accuracy of billing data in an academically affiliated family practice.

Methods. The progress notes from 1253 consecutive visits were independently reviewed by two family physicians, and the diagnoses, use of procedures, and level of service were determined for each visit. Discrepancies between the reviewers were resolved by consensus. These data were compared with the data on the corresponding billing form that had been completed by the care providers (ie, physicians on the faculty, physicians in training, family nurse practitioners, and nurses).

Results. There was poor agreement between the billing form and progress note on level of service and number of diagnoses ($\kappa = 0.37$ and $\kappa = 0.28$, respectively). The progress note usually indicated that a higher level of service should have been billed for a visit than actually was billed. Underreporting of the number of diagnoses was substantial; the billing forms listed only 69% of the diagnoses identified in the progress notes. In 60% of visits, each diagnosis on the billing form had a matching diagnosis in the progress note. This could be improved to 78% of visits if broad categories of disease were used. Residents were similar to faculty in the accuracy of reported level of service and types of diagnoses, but were more likely to underreport the number of diagnoses.

Conclusions. Ambulatory care data from computerized billing files may not be sufficiently accurate for proper reimbursement of physician services or for use in research.

Key words. Fees and charges; documentation; practice management, medical. *J Fam Pract 1991; 33:593-598.*

For more than two decades, the federal government has been concerned with rising health care costs. In 1983, the Health Care Financing Administration sought to control hospital costs by making diagnostic and procedural data the primary determinant of hospital reimbursement through the prospective payment system. Recently, the US Congress has implemented a similar approach to control physician costs in the ambulatory care setting for the Medicare program. Through the Omnibus Budget and Reconciliation Act of 1989, any office procedure provided to Medicare patients and the reason (ie, diagnosis) must be reported in a standard format. From these data, Medicare authorities determine medical necessity and "appropriate" level of reimbursement; payment may be denied or reduced for diagnostic and office procedural services that do not fall within the defined limits for the given diagnosis.

The hospital discharge database and its ambulatory care analogue, the patient encounter database, are the sources of diagnostic and procedural information for patient and third-party billing. These databases also are used for health services research, such as the assessment of quality of care.[1–7] Consequently, the reliability and validity of the data are essential for appropriate reimbursement and valid research.

The reliability and validity of hospital discharge data have been investigated since the 1970s. Reports consistently demonstrate that inpatient data are inaccurate, particularly for diagnoses.[1–6] Some researchers believe sufficient error exists in the diagnostic information to render hospital discharge data inadequate for "detailed research and evaluation."[3(p1003)]

Few similar evaluations of the reliability and validity of the ambulatory care database have been conducted, and the results vary.[8–12] Level of agreement between diagnoses listed in the office medical record and the

Submitted, revised, September 4, 1991.

From the Department of Family Medicine, East Carolina University School of Medicine, Greenville, North Carolina. Requests for reprints should be addressed to Ronnie D. Horner, PhD, HSR&D (152), VA Medical Center, 508 Fulton St, Durham, NC 27705.

ISSN 0094-3509

observed events of the visit range from 60% to 90%.[8–10] Studies of the quality of computerized medical files, including patient billing files, show that between 30% and 50% of visits contain errors in the recorded data.[11,12] Underreporting of diagnoses is the major type of error; coding errors are a minor problem.[10–12] These reports are based on relatively few visits, however, ie, 26 to 150 visits; hence, they may not reflect the actual degree of validity and reliability of billing files.

This study presents the results of an investigation into the accuracy of billing information in a family practice, and is based on a large number of visits. The recording behavior of physicians on the faculty, physicians in training, family nurse practitioners (FNPs), and registered nurses was examined.

Methods

The data for this study were from an academically affiliated family practice in eastern North Carolina, which is the training site for a residency program in family medicine. The medical staff consisted of 36 residents and a faculty of 12 family physicians. This practice had approximately 21,000 active patients who accounted for about 36,000 visits in 1988. The patients were from the city in which the practice was located (population 45,000) and from the surrounding rural communities, and represented all sociodemographic groups.

All encounters during the 2-week period, August 24 through September 7, 1988, constituted the database. This period was selected without regard to either number or type of visits, but with the intent that the period be representative with regard to billing information.

Patient billing was based on the encounter form, which specified the level of service, up to five diagnoses, and the types of procedures or laboratory tests ordered. The patient's age, sex, race, method of payment, and name of provider were also indicated. These data were abstracted from each encounter form, which is referred to as the "billing form" in this article.

The progress note in the medical record that corresponded to each visit was reviewed without reference to the billing form data to determine the "standard" level of service, diagnoses, and procedures. (Previous studies show that information on the chief complaint and "review of systems" in the medical record is fairly reliable and valid, ranging from 70% to 90% concordance with events of the actual visit.[8,9]) The initial review was performed independently by two family physicians (J.R.P. and F.H.L.). The results from the two reviewers were compared and discrepant results resolved by consensus; disagreement occurred for approximately one third of the study encounters, but the discrepancies were minor and almost always involved level of service rather than diagnoses. These data are referred to as the "billing standard" in this report.

Level of service, diagnoses, and procedures were classified according to standard definitions. The level of service was based on the Current Procedural Terminology (CPT) codes for office visits[13]; level of service is independent of diagnosis and is based on elements of the history and physical examination, complexity of the case, and management plan. Diagnoses of the standard were in the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) format. Diagnoses on the billing form were coded in the format of the International Classification of Health Problems in Primary Care (ICHPPC-2); these diagnostic codes were converted to codes of the ICD-9-CM, based on the comparability of the codes as presented in the ICHPPC-2 manual.[14] Categories of related diagnoses were created based on the major groupings of disease (eg, infectious and parasitic diseases) used in the ICHPPC-2 and ICD-9-CM coding schemes. Procedures on the billing form were coded according to CPT codes. Because of time constraints, however, the reviewers only decided if a procedure should have been coded for the visit (ie, a dichotomous variable); thus, the billing standard did not specify the type of procedure.

Data analysis involved summarizing the information by percentages and means (± standard deviation). The kappa statistic and 95% confidence intervals (95% CI) were used to specify the degree of agreement between the billing form and the billing standard.[15,16] The methods described by Fleiss[16] were used to make statistical comparison of the kappas among types of providers.

Results

Encounter Characteristics

During the 2-week period, 1140 patients visited the clinic. Most patients (90.1%) made only one visit during the study period; 8.7% made two visits, and 1.2% made three or more visits. Four patients came for laboratory studies only and were not considered further. The characteristics of the remaining 1136 patients are presented in Table 1. Ninety-five percent of the visits involved diagnosis and treatment without any procedures, while 2.5% involved an office-based procedure only; the remaining visits (2.5%) involved diagnosis, treatment, and a procedure. A faculty member or third-year resident provided care in 67.5% of all visits, a family nurse practitioner (FNP) was seen in 11.5% of visits, and a second-

Table 1. Characteristics of Patients Who Visited the Family Practice Clinic During the Two-Week Study Period (N = 1136)*

Characteristic	No. of Patients (%)
Sex	
Male	397 (34.9)
Female	739 (65.1)
Race	
White	625 (55.0)
Black	511 (45.0)
Age (in years)	
0–14	122 (10.7)
15–24	135 (11.9)
25–34	244 (21.5)
35–44	231 (20.3)
45–54	131 (11.5)
55–64	112 (9.9)
65 and older	161 (14.2)
Type of insurance	
Self-pay (none)	200 (17.6)
Medicaid	195 (17.2)
Medicare	124 (10.9)
HMO	319 (28.1)
Commercial carrier	298 (26.2)
No. of visits (during study period)	
One	1024 (90.1)
Two	99 (8.7)
Three or more	13 (1.2)

**Excludes four patients who came for laboratory work only.*
HMO denotes health maintenance organization.

year resident in 15% of visits. First-year residents were the providers in approximately 3% of visits, as were nurses.

Level of Service

Table 2 shows the degree of concordance between the billing form and the billing standard on level of service. Agreement was poor ($\kappa = 0.37$, 95% CI = 0.33 to 0.40) with the billing form generally showing a lower level of service than indicated by the billing standard. Degree of agreement for level of service did not differ significantly among types of providers, although the more experienced providers had higher kappas (Table 3). Patient demographic characteristics and type of health insurance were not associated with degree of agreement on level of service.

Errors of Omission

The second issue examined was the extent of omission of diagnoses in the computerized billing database (Table 4). Again, agreement was poor between the number of diagnoses recorded on the billing form and the number of diagnoses listed on the billing standard for the visit ($\kappa = 0.28$, 95% CI = 0.24 to 0.31).

Where agreement was lacking, the billing form usually had too few diagnoses listed. For 33 visits (2.6% of all visits), the billing form contained more diagnoses than the billing standard, whereas in 525 visits (41.9% of all

Table 2. Concordance Between the Billing Form and the Billing Standard in Level of Service per Visit (Number of Visits)

Level of Service on the Billing Form	Level of Service on the Billing Standard								
	No Charge	Nurse Visit	Routine	Intermediate	Extended	Comprehensive	Obstetrics Care	Procedure Only	Procedure and Cognitive Services
No charge	12	10	5	3	1	0	0	4	1
Nurse visit	0	11	1	1	0	0	0	0	0
Routine	16	0	269	211	28	1	7	1	8
Intermediate	0	0	77	278	85	9	7	6	11
Extended	0	0	1	25	27	7	0	1	1
Comprehensive	0	0	0	3	7	11	2	1	1
Obstetrics care	0	0	2	1	0	0	51	0	0
Procedure only	2	0	0	0	0	0	0	18	0
Procedure and cognitive services	0	0	2	1	1	0	0	2	10

κ (observed level of agreement-expected level of agreement)/(1 − expected level of agreement) = 0.37. (Observed agreement = proportion of times the billing form and the billing standard agreed on level of service; expected agreement = proportion of times the billing form and the billing standard are expected to agree by change alone on level of service.)
NOTE: Billing standard *refers to the level of service, diagnoses, and procedures, obtained by review of the patient's progress note.*

Table 3. Level of Agreement Between the Billing Form and the Billing Standard in Level of Visit, by Type of Provider

Type of Provider	No. of Visits	Level of Agreement (κ)
All	1238	0.36
Residents		
1st year	33	0.19
2nd year	185	0.28
3rd year	400	0.33
Faculty	446	0.43
Family nurse practitioners	143	0.28
Nurses	35	0.32

Test of differences among providers in level of agreement (kappa): $\chi^2 = 6.32$, P = 0.29.
NOTE: Billing standard *refers to the level of service, diagnoses, and procedures, obtained by review of the patient's progress note.*

visits), the billing standard had the greater number of listed diagnoses. The patient's age, race, sex, and type of insurance did not influence the degree of agreement for number of diagnoses.

First-year residents had the worst level of agreement between the number of diagnoses listed on the billing form and the billing standard, with a kappa of 0.15 (Table 5). Nurses had the best level of agreement with a kappa of 0.48; however, 97% of visits to nurses involved only a single diagnosis. The faculty physicians, FNPs, and second- and third-year residents had kappas ranging from 0.20 and 0.34, indicating poor agreement beyond chance in the number of diagnoses recorded on the billing form and those from the billing standard. Residents had significantly poorer agreement between number of diagnoses on the billing form and the billing standard ($P < .001$) than other providers (ie, physicians on the faculty, FNPs, and nurses).

Table 4. Concordance Between the Billing Form and the Billing Standard in Number of Diagnoses per Visit* (Number of Visits)

No. of Diagnoses on the Billing Form	No. of Diagnoses on the Billing Standard					
	1	2	3	4	5	Total
1	518	232	97	40	14	901
2	18	142	74	23	12	269
3	2	7	30	16	13	68
4	1	1	4	3	4	13
5	0	0	0	0	2	2
Total	539	382	205	82	45	1253

**Fourteen visits were excluded because no diagnoses were listed on the billing form.*
κ (observed level of agreement-expected level of agreement)/(1−expected level of agreement) = 0.28.
(Observed agreement = proportion of times the billing form and the billing standard agreed on number of diagnoses; expected agreement = proportion of times the billing form and the billing standard are expected to agree by chance alone on number of diagnoses.)
NOTE: Billing standard *refers to the level of service, diagnoses, and procedures, obtained by review of the patient's progress note.*

Table 5. Level of Agreement Between the Billing Form and the Billing Standard in Number of Diagnoses per Visit, by Type of Provider

Type of Provider	No. of Visits	Level of Agreement (κ)
All	1253	0.28
Residents		
First year	36	0.15
Second year	192	0.20
Third year	398	0.22
Faculty	448	0.33
Family nurse practitioners	144	0.34
Nurses	35	0.48

Test of differences among providers in level of agreement (kappa): $\chi^2 = 35.48$, P<.001.
NOTE: Billing standard *refers to the level of service, diagnoses, and procedures, obtained by review of the patient's progress note.*

Errors of Coding

The third issue examined was the degree of diagnostic accuracy, that is, the extent to which diagnoses recorded on the billing form matched the diagnoses listed on the billing standard for the visit. Although the list of diagnoses on the billing form tended to be an undercount, all of the diagnoses listed on the billing form matched a diagnosis from the billing standard in only 60% of visits. For 27% of all visits, none of the diagnoses on the billing form matched a diagnosis on the billing standard. Physicians on the faculty, first-, second-, and third-year residents, and FNPs had similar percentages of patient visits for which the billing form diagnoses matched the diagnoses on the billing standard exactly; nurses had only half the percentage of visits with perfect matches for diagnoses as the other providers. Furthermore, listed diagnoses on 63% of nurses' billing forms did not match the billing standard, despite the fact that most of these visits had a single diagnosis on the billing form. For FNPs, faculty, and third-year residents, one fourth of their encounters had no diagnostic matches. Twenty percent of patient visits to first- and second-year residents did not match any diagnoses.

The use of major ICD-9-CM diagnostic categories rather than specific diagnoses yielded an additional 10% to 15% of visits in which all diagnoses on the billing form had a match in the billing standard. Thus, in 78% of visits, all of the billing form diagnoses had a matching

diagnosis on the billing standard. Performance of the various types of providers was similar to the exact-matching situation.

For visits where one or more billing form diagnoses had a matching diagnosis in the billing standard, the degree of matching decreased with the order of the diagnosis. Thus, for visits with matches, 80% of the first billing form diagnosis matched exactly with a diagnosis in the billing standard. The percentage of matched diagnoses dropped to 19% for the second diagnosis on the billing form, and even fewer third and fourth billing form diagnoses could be matched.

The greatest inaccuracies on the billing forms were found for endocrine diseases; mental disorders; supplemental conditions such as preventive medicine and social problems; and symptoms, signs, and ill-defined conditions. Each of these four major diagnostic categories accounted for approximately 10% to 15% of all of the omitted or incorrectly specified diagnoses on the billing form, and together represented 45% of all such errors. Most of the other major ICD-9-CM diagnostic categories accounted for 5% to 7% of omitted or incorrectly identified diagnoses.

Discussion

The patient billing database and its inpatient analogue, the hospital discharge database, are crucial to reimbursement for health care services and have the potential to facilitate health services research, including investigation of quality of care issues. Although the poor quality of the diagnostic data in the hospital discharge database has been documented consistently,[1–6] previous studies of the ambulatory care database suggest that these databases may be of better quality than those of the hospital.[10–12]

This investigation found, however, that the quality of the ambulatory care billing database is generally poor. There was a low degree of agreement between the level of service, the number of diagnoses, the specified diagnoses in the billing database, and the content of the visit as described in the progress note. The billing file showed a lower level of service and too few diagnoses; and, for almost one fourth of visits, the listed diagnoses were not found in the progress note. Although the use of major diagnostic categories resulted in 78% of visits having all listed diagnoses present in the progress note, the problem of too few diagnoses being listed remains. Moreover, broad categories of disease may lack the specificity required for research. Although providers other than residents had a higher degree of completeness or accuracy, their billing form data still contained substantial errors.

Explanation for the diagnostic inaccuracy may relate to the billing form per se. The codes of common diagnoses, listed on the back of the form for reference by the provider, were organized numerically by ICHPPC-2 codes. Thus, the provider had to "screen" sets of codes to find the appropriate diagnosis, look up the appropriate code in the coding manual, or recall the code from memory. Failure to record all diagnoses may be related to the time constraints within the patient care setting or the provider's failure to appreciate the importance of diagnostic completeness on the billing form. The tendency to record a lower level of service may reflect poor understanding of the true value of the physician's time and cognitive services.

Methods of improving coding accuracy have been proposed.[17] One solution to the problem of billing form errors may be instruction of residents on the importance of accuracy in reporting level of service and diagnoses. Such a program was instituted at the family practice for this investigation. Informal review of subsequent billing data indicated more accurate reporting of the level of service.

We recognize that this study of a single, academically affiliated family practice may not represent the situation in private family practices or in other primary care specialties. It has been suggested that teaching centers are notorious for poor data quality.[10] If that is the case, then our findings may be liberal in the degree of inaccuracy revealed. Since physicians receive their initial training in practice management at these centers, however, it seems reasonable to expect the learned behaviors to continue in a different practice setting, at least in part.

If validated in other studies, these results have important implications for current reimbursement issues. Reimbursement for ambulatory care services is based on patient billing form data. Whether a commercial insurance carrier (such as Blue Cross/Blue Shield) reimburses a provider, and the amount of that reimbursement, depends on the reported diagnoses. Inaccurate diagnoses may result in denial of the claim, leaving the physicians to either (attempt to) collect from the patient or receive no reimbursement at all. Payments from the Medicare program, where only a portion of established charges are reimbursed, also could be affected. For example, billing for a procedure without an appropriate corresponding diagnosis may result in denial of the charge by Medicare authorities.

A consequence of inaccurate patient billing data is that primary care physicians may be systematically reducing their income. The degree of this reduction in income is being explored. It may be that improved accuracy in patient billing data will do more to remedy the income situation of primary care physicians than the much ad-

vocated changes in reimbursement for "cognitive services."[18]

For research purposes, the poor quality of the diagnostic data casts doubt on the results of previous studies that have relied on the billing database, and call into question the usefulness of future studies that may use these data. For example, the numerous studies of family practice and primary care diagnoses may have provided erroneous descriptions of content. Methods for classifying problems that override "idiosyncratic labeling and coding habits of individual physicians" cannot adjust for the substantial underreporting of diagnoses.[19] Moreover, unless major categories of diagnoses are used, the content of care may be misleading. As with its inpatient analogue, the ambulatory care billing database may be too imprecise in diagnostic content to be useful for health services research.

References

1. Institute of Medicine. Reliability of hospital discharge abstracts. Washington, DC: National Academy of Sciences, February 1977.
2. Institute of Medicine. Reliability of Medicare discharge abstracts. Washington, DC: National Academy of Sciences, November 1977.
3. Demlo LK, Campbell PM, Brown SS. Reliability of information abstracted from patients' medical records. Med Care 1978; 16: 995–1005.
4. Institute of Medicine. Reliability of national hospital discharge survey data. Washington, DC: National Academy of Sciences, 1980.
5. Roos LL, Roos N, Cageorge SM, Nicol JP. How good are the data? Reliability of one health care data bank. Med Care 1982; 20:266–76.
6. Lloyd SS, Rissing JP. Physician and coding errors in patient records. JAMA 1985; 254:1330–6.
7. Anderson JE. Reliability of morbidity data in family practice. J Fam Pract 1980; 10:677–83.
8. Romm FJ, Putnam SM. The validity of the medical record. Med Care 1981; 19:310–5.
9. Moran MT, Wiser TH, Nanda J, Gross H. Measuring medical residents' chart-documentation practices. J Med Educ 1988; 63: 859–65.
10. Bentsen BG. The accuracy of recording patient problems in family practice. J Med Educ 1976; 51:311–6.
11. Gehlbach SH. Comparing methods of data collection in an academic ambulatory practice. J Med Educ 1979; 54:730–2.
12. Fortinsky RH, Gutman JD. A two-phase study of the reliability of computerized morbidity data. J Fam Pract 1981; 13:229–35.
13. American Medical Association. Physicians' current procedural terminology. Chicago: American Medical Association, 1984.
14. WONCA. ICHPPC-2-defined (International Classification of Health Problems in Primary Care). 3rd ed. Oxford: Oxford University Press, 1983.
15. Agresti A. Categorical data analysis. New York: John Wiley & Sons, 1990:365–70.
16. Fleiss J. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, 1981:217–25.
17. Waterstraat FL. Diagnostic coding quality and its impact on health care reimbursement. J Am Med Rec Assoc 1990; 61:52–9.
18. Lee PR, Ginsburg PB, Leroy LB, Hammons GT. The physician payment review commission report to Congress. JAMA 1989; 261:2382–5.
19. Schneeweiss R. Diagnosis clusters: a new approach for reporting the diagnostic content of family practice residents' ambulatory experiences. J Fam Pract 1985; 20:487–92.