A Comparison of Ambulatory Medicaid Claims to Medical Records: A Reliability Assessment

Donald M. Steinwachs, Ph.D.,* Mary E. Stuart, Sc.D.,† Sarah Scholle, Dr.P.H.,‡ Barbara Starfield, M.D.,* Michael H. Fox, Sc.D.,§ and Jonathan P. Weiner, Dr.P.H.*

*Department of Health Policy and Management, Johns Hopkins University, School of Hygiene and Public Health, Baltimore, Maryland; †University of Maryland at Baltimore County, Maryland; ‡University of Pittsburgh, Pittsburgh, Pennsylvania; and §University of Kansas, Lawrence, Kansas

This study compares the documentation of ambulatory care visits and diagnoses in Medicaid paid claims and in medical records. Data were obtained from Maryland Medicaid’s 1988 paid claims files for 2407 individuals who were continuously enrolled for the fiscal year, had at least one billed visit for one of six indicator conditions, and had received the majority of their care from one provider. The patients sampled were also stratified on the basis of the case-mix adjusted cost of their usual source of care. The medical records for these individuals as maintained by their usual source of care were abstracted by trained nurse reviewers to compare claims and record information. Linked claim and medical record data for sampled patients were used to calculate: (i) the percent of billed visits documented in the record, (ii) the percent of medical record visits where both the date and the diagnosis agreed with the claims data, and (iii) the ratio of medical record visits to visits from billed claims. Included in the analysis were independent variables specifying place of residence, type and costliness of usual care source, level of patient utilization, and indicator condition on which patient was sampled. Ninety percent of the visits chronicled in the paid claims were documented in the medical record with 82% agreeing on both date and diagnosis. Compared to the medical records kept by private physicians and community health centers, a significantly lower percent of hospital medical records agreed with the claims data. Total volume of visits was 2.6% higher in the medical records than in the claims. Claims data substantially understated visits in the medical record by 25% for low cost providers and by 41% for patients with low use rates (based on claims information). Conversely, medical records substantially understated billed visits by 19% for rural patients and by 10% for persons with high visit rates. Although Medicaid claims are relatively accurate and useful for examining average ambulatory use patterns, they are subject to significant biases when comparing subgroups of providers classified by case-mix adjusted cost and patients classified by utilization rates. Medicaid programs are using claims data for profiling and performance assessment need to understand the limitations of administrative data.

INTRODUCTION

Medicaid Management Information Systems (MMIS) were developed in the 1980s to meet basic requirements for paying claims, accounting for state and federal matching funds, and monitoring fraud and abuse. As with other administrative health information systems (e.g., Medicare, private insurers), there are new demands for expanded applications. These include monitoring the quality of preventive, acute, and chronic care; profiling provider cost and quality performance; evaluating policy; and conducting research. There is concern that the quality of claims data (i.e., completeness and accuracy) cannot support these expanded applications (1–3). This work reports on a comparison of Maryland Medicaid claims and patient medical record data from private office practices, federally qualified health centers (FQHC), and hospital outpatient settings to assess the reliability of claims data.

BACKGROUND

There has been growing interest in the application of administrative data for policy evaluation, provider
These include diagnoses (particularly where there are more complete and accurate than the claims record. Where the medical record would be expected to be likely to contain a claim for any prior mammogram beformed before hospitalization, a Medicaid claims file is the latter as the standard. However, previous studies have pointed to the potential of administrative data for examining quality of care (7, 8).

The strengths of administrative databases are that they generally provide comprehensive information for an insured population and are available with no additional data collection costs. Limitations include inadequate specification of clinical detail and uncertainties regarding the reliability of the information, especially diagnosis information (9). Since the introduction of Diagnosis Related Groups for payment under Medicare, providers have given more attention to the coding of inpatient discharge diagnoses. So far nothing of a similar nature has occurred to encourage providers to focus greater attention on the coding of outpatient diagnoses.

Most studies using insurance claims data have focused on inpatient care. For example, Wennberg and colleagues (5) examined variations in admission and procedure rates and the relationship of these variations to death rates. Roos et al. (8) and Iezzoni et al. (10), used administrative data on hospital readmissions to examine variations in the rate of surgical complications, including death. In general, these studies found that inpatient diagnosis and procedure information were reasonably reliable.

In contrast, relatively few studies have used outpatient diagnostic information from claims. The early studies relied on data collected through HMO administrative systems or from medical record abstracts (11, 12). Not until the 1980s did private insurers and state Medicaid agencies begin to require ICD-9 coded diagnoses for outpatient services. It was only in 1991 that Medicare required that physicians record a diagnosis on outpatient claims. As a result, only recently has information on the reliability of outpatient diagnostic information (9).

The traditional approach for testing the validity of claims data is to compare claims data with medical records using the latter as the standard. However, previous studies have suggested that some types of information may be more reliably captured through claims data than through medical records. For example, although a hospital medical record may be silent on whether a mammogram was performed before hospitalization, a Medicaid claims file is likely to contain a claim for any prior mammogram because all providers bill Medicaid (13).

In contrast, there are categories of information where the medical record would be expected to be more complete and accurate than the claims record. These include diagnoses (particularly where there are multiple diagnoses), medical histories, results of laboratory tests or procedures, and non-billable therapies. It is clear that comparisons between Medicaid claims data and the patient’s medical record can provide information on the completeness of both of these alternative sources of information.

METHODS

This study comparing ambulatory Medicaid claims and medical records was undertaken as part of a larger study that assessed geographic variations in ambulatory service utilization and the relationship of quality of care to costs (14–16). The study population consisted of adults and children who were continuously enrolled in the Maryland Medicaid program under Aid for Families with Dependent Children (AFDC), Supplemental Security Income (SSI), or General Public Assistance (supported by state funds only) throughout fiscal year 1988 (July 1, 1987, through June 30, 1988); resided in specifically defined urban, suburban, and rural areas of Maryland; and, based on claims information, had one of the following six indicator conditions: asthma, diabetes, otitis media, hypertension, pregnancy, well child. For three of these indicators (hypertension, asthma, diabetes) eligibility of an enrollee for the study required a billed visit with the indicator diagnosis once during the study year (fiscal year 1988) or in the preceding year. For otitis media, the patient had to have at least one episode of care that started during the year. To examine well-child care, we selected children who had their second birthday during the first half of the study year. To ensure that we would have the full set of claims for prenatal care, we selected patients for the pregnancy sample based on an inpatient delivery claim during the last quarter of the year, March 1 to June 30, 1988. We identified visits using Current Procedural Terminology (CPT) Codes (90000–90700, 90750–90781, 90800–90899, and 59400–59599) as well as special Medicaid billing codes used for EPSDT visits, local health department visits, and hospital clinical services.

Patients were linked to providers through a process that identified the patient’s usual care source based on utilization during the study year. The usual care source was defined as the provider who billed more than half of all the patient’s ambulatory visits during the year. Some Medicaid enrollees could not be assigned a usual care source either because they had no ambulatory visits during the year or because their visits were distributed such that no single provider had received a majority of the visits. Individuals without a usual care source and individuals enrolled in HMOs, where no claims data were
available, were not included in the sample. In addition, for an enrollee to be included the enrollee’s usual source of care had to serve as the usual source of care for at least nine other enrollees in the study population. These usual source of care providers were stratified by organizational category and included 106 private physicians, 10 FQHCs, and 19 hospitals with outpatient services. The private physician category included group practices as well as solo and partnership practices. Most hospitals included in the analysis had outpatient departments and stand-alone clinics. In some cases the usual care source for a patient was the emergency room of a hospital where there were no other outpatient clinics.

Within these provider practices, patients were sampled on the basis of indicator condition as well as for the relative costliness of their usual source of care. For each provider, the case-mix adjusted cost of all Medicaid patients who had used that provider as a usual source was calculated using Ambulatory Care Groups (17, 18). The distribution of case-mix adjusted cost by providers within each organizational type was then calculated. Those in the top thirty percent were classified as “high cost”; those in the bottom thirty percent, as “low cost”; and those in the middle forty percent, as moderate cost. The classification was done within organizational type because the Maryland Medicaid Program reimburses hospitals, FQHCs, and private physicians differently for the same services (19).

In addition, patients were classified by their relative level of utilization. Utilization was determined by taking into account not just those visits with the sampled diagnosis but all visits made by the patients to their usual care source during the fiscal year. Patients were also classified as to whether they resided in Baltimore City; in a suburban county, that is, in the remainder of the Baltimore metropolitan area; or in a rural Eastern Shore county.

A total of 2407 adults and children were sampled. In Table 1, their characteristics are shown by place of residence, organizational type of usual source of care, costliness of usual source of care, level of utilization, and ICD-9-CM indicator diagnoses.

The medical records maintained by the usual source of care for each of these sampled patients were abstracted by the Delmarva Foundation for Medical Care (a federally designated Peer Review Organization). Providers were notified of the study through a letter sent to them by the Maryland Medicaid Compliance Administration. Each provider’s letter included a list of the patients to be reviewed and requested that all charts relating to their care for a 2-year period be made available for review in the provider’s office by nurse chart abstractors. Abstractors were trained by the research staff. Supervisors monitored their abstracting performance over the course of the data collection. Interrater reliability, based on reabstracting a sample of charts indicated agreement of over 90% for the items reported on here.

For each sampled patient a “face sheet” was generated from the Medicaid paid claims file. The face sheet included the date and diagnosis of each visit billed by the patient’s usual source of care from July 1, 1987, through June 30, 1988. Using the face sheet, the abstractor marked whether there was evidence in the chart of an encounter on the exact date billed and whether the billed diagnosis was mentioned. If any other diagnoses were noted in the chart, the abstractor recorded up to three of them. The abstractor also recorded any visits in the record that were not billed, along with the diagnoses for those visits.

There was a 100% participation of providers in chart review. Overall, 92% of the sampled patient records were located and abstracted. Of those found, 91% were eligible for review. To reach the target for the number of record reviews in each provider, additional patients were added from a substitute list for each provider.

### Table 1

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>1747</td>
<td>73</td>
</tr>
<tr>
<td>Suburban</td>
<td>421</td>
<td>17</td>
</tr>
<tr>
<td>Rural</td>
<td>239</td>
<td>10</td>
</tr>
<tr>
<td>Usual source of care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private physician</td>
<td>1095</td>
<td>45</td>
</tr>
<tr>
<td>Community health center</td>
<td>270</td>
<td>11</td>
</tr>
<tr>
<td>Hospital</td>
<td>1042</td>
<td>43</td>
</tr>
<tr>
<td>Costliness of usual source of care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1004</td>
<td>42</td>
</tr>
<tr>
<td>Moderate</td>
<td>573</td>
<td>24</td>
</tr>
<tr>
<td>Low</td>
<td>830</td>
<td>34</td>
</tr>
<tr>
<td>Patient utilization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (10+ visits)</td>
<td>694</td>
<td>29</td>
</tr>
<tr>
<td>Moderate (4–9 visits)</td>
<td>1082</td>
<td>45</td>
</tr>
<tr>
<td>Low (1–3 visits)</td>
<td>631</td>
<td>26</td>
</tr>
<tr>
<td>Indicator condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>393</td>
<td>16</td>
</tr>
<tr>
<td>Diabetes</td>
<td>368</td>
<td>15</td>
</tr>
<tr>
<td>Otitis media</td>
<td>464</td>
<td>19</td>
</tr>
<tr>
<td>Hypertension</td>
<td>592</td>
<td>28</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>384</td>
<td>16</td>
</tr>
<tr>
<td>Well child</td>
<td>206</td>
<td>09</td>
</tr>
</tbody>
</table>

### Analysis

The analysis of Medicaid claims reliability focuses on four questions:
Is the visit shown on the claim documented in the medical record? A visit appearing on a billed claim is documented if on the same day there is documentation of a visit in the medical record (chart). The percent of billed visits documented in the chart was calculated by dividing the number of visits found in the chart by the number of visits billed for the same day in the claims.

**For billed visits, is the diagnosis listed on the claim in the medical record?** This is examined using the percent of billed visits having both a matching date and diagnosis in the patient's record. Diagnosis agreement is based on 3-digit ICD-9 codes. The numerator is the number of billed visits with both matching dates and diagnoses which are in the medical record and the denominator is the number of visits billed for the same day.

How do the volume of visits identified in the medical record compare with the volume of visits billed? A volume ratio is calculated by dividing the number of visits in the chart by the number of visits in the claims file.

What characteristics of patients and of their usual sources of care are associated with higher and lower agreement between the chart and the bill? Common applications of claims data for profiling providers and patients include identifying high cost versus low cost providers after adjusting for case-mix and identifying high utilization and high cost patients. Regression models were used to examine percent of agreement on diagnosis and date and the ratio of visits in the chart and claims data, adjusting for sample design.

## RESULTS

### Was the Billed Visit Documented in the Medical Record?

Overall, 89.7% of billed visits were documented in the record with coinciding dates (Table 2). This varied by type of usual source of care. Visits by patients whose usual source of care was a FQHC had the highest documentation rate (94.6%), visits by patients of private physicians were next highest (93.1%), and visits by patients of hospitals were lowest (86.1%).

To examine the possibility that the source of error related to the date of the visit, documentation rates were calculated for the samples of patients with hypertension, diabetes, and pregnancy using a 7-day time window around the billing date. One-third to one-half of the 10% of visits not documented in the chart on the same date as shown on the claim were found to correspond to visits on other dates within the seven day time period.

### For Billed Visits, Was the Diagnosis Listed on the Claim in the Medical Record?

Table 2 also shows that for 81.8% of all visits billed, both the exact date and the diagnosis were documented in the medical record (Table 2). This percentage varied from a low of 76.1% for hospitals to a high of 86.5% for FQHCs. This estimate may be biased toward the low side because we limited the medical record abstractors to three diagnoses. Medicaid bills had one diagnosis listed in 78% of the sample and two diagnoses in 21%, which was the maximum allowable. By comparison, 30% of the visits documented in the medical record had two or more diagnoses, and 8% had three or more diagnoses listed.

### How Did the Volume of Visits Identified in the Medical Record Compare with the Volume of Visits Billed?

Another issue that has been raised regarding Medicaid claims data is that reimbursement levels may be so low that some providers may not bother to bill for visits. This study could not identify providers who never submitted a bill for any patient, but it was possible to examine the numbers of visits in the record and their relationship to
the numbers of visits billed. The average ratio of visits listed in the chart to those billed was 1.026, i.e., 2.6% higher in the chart. This ratio was close to one (0.993) in hospitals and was 9.4% higher in FQHCs and 4.2% higher among private physicians. These findings suggest there could be from 4–9% underbilling in FQHCs and private physician offices.

**What Characteristics of Patients and of Their Usual Sources of Care Are Associated with Higher and Lower Agreement between the Chart and the Bill?**

Understanding average differences is important, but this does assure that claims data can be used reliably to identify high versus low cost providers, or to identify patients with low or high use patterns in specific diagnostic categories.

In Table 3, a least squares regression model is shown of the percent of agreement on diagnosis and date of billed claims to the patient's medical record. The findings show consistent differences by type of usual care source, with a statistically higher match of same date and diagnosis for private physicians than for hospitals. In addition, statistically significant and small differences are present by costliness of the usual source of care (3% lower agreement among low cost than moderate cost usual sources of care). Among low use patients there is 3% lower agreement than moderate use, and there is 2% higher agreement among high use patients. Patient's place of residence is significant with 4% higher agreement for rural residences over urban residents. Also, there is some variation among the indicator conditions (i.e., 4% less agreement for persons with diabetes than the reference category, hypertension).

To examine visit volume differences, Table 4 shows regression results for the ratio of chart visits to billed claims for the usual source of care. Overall, there were 2.6% more visits in the medical record than there were Medicaid claims from the usual care source. Although this is a relatively small average difference, patients of low cost providers had almost 25% more visits recorded in their provider's medical record than billed visits in the claims data. This suggests that one reason a usual source of care may appear to be low cost is that it has not billed for a substantial proportion of patient visits.

In some quality of care applications the focus is on under-use or over-use of services. As a group, the 631 patients classified as "low use," had 41% more visits recorded in the medical record than in the Medicaid claims. In contrast, the high-use group of 694 patients had 10% fewer visits recorded in the medical record than in the claims. The low use group had substantial missing
claims data, whereas the high use group lacked chart documentation for 10% of visits billed.

Another group for which discrepancies existed was the 239 patients classified as rural, where 19% of visits were not documented in the medical record. One caveat, these findings are based on relatively few providers. A comparison of the two regression models indicates that rural patients had better correspondence between the claims and charts for visits that were billed, but that they also had more billed visits undocumented in the medical record. Also, the combined results suggest that the billing errors more common among low cost providers and low use patients are related to both poorer agreement for billed visits and fewer total visits being billed.

DISCUSSION

There are several factors which limit the generalizability of the findings. The study dealt with only one state Medicaid system and spanned a single year. The rural patients were limited to one geographic area of the state and a relatively small number of usual sources of care. The usual source of care predictably included mainly primary care providers; the study had little ability to assess correspondence for specialists. Only patients with at least one ambulatory visit for one of the six indicator conditions were eligible for the study, and patients had to have the majority of their visits with a single provider. In addition, providers had to have at least ten Medicaid patients in their practice who were continuously enrolled for the year and for whom they were the usual source of care.

The findings from this detailed comparison of Medicaid claims for ambulatory care visits to the medical records of the usual care source are somewhat encouraging. On average, 90% of the billed visits were documented in the medical record. Apparent differences in the dates recorded on the bill and in the medical record explained about one-third of these discrepancies but left the remaining two-thirds unexplained.

In terms of the volume of visits, roughly offsetting differences were found. Approximately 10% of billed visits did not have a matching visit on the same date in the medical record and 12–13% of visits in the medical record did not have a billed claim for the same date. The overall result was 2.6% more visits in the record than there were Medicaid claims. One reason for visits not being billed might occur when the patient was seen by a non-physician provider and the service rendered may not be billable without physician involvement. We do not know how frequently this occurred but expect it would have been relatively rare.

The findings show lower agreement matching both date and diagnosis from the claim to the medical record, with an overall rate of 82% agreement. The rates of agreement were lowest among hospitals, at an average of 75%. One would expect that routine administrative systems should do be able to do substantially better than the rates of agreement found.

The findings raise some concerns about the application of claims data for identifying high versus low cost providers and identifying low use patients. Providers classified as low cost based on case-mix adjusted claims experience for their patients had 25% more visits chronicled in the medical record than in the claims. Even though these providers appeared low cost based on billed claims, their actual practice patterns may not have been low cost if they had submitted claims for all services documented in the patient's medical record. Similarly, patients with low visit rates based on claims data had 40% more visits in their medical records than there were bills.

Instances where there were more billed visits on average than found in the medical record occurred primarily among patients with a high volume of visits (10% fewer visits in the medical record than billed) and among rural patients (19% fewer visits in the medical record than billed). It is unknown whether these discrepancies indicate a serious documentation problem, erroneous billing, or a combination of both.

CONCLUSION

In summary, the findings are very positive for using Medicaid claims for examining averages. However, the biases become more significant when examining distributional characteristics. For patients with low utilization rates it appears that the Medicaid claims data may be substantially incomplete, whereas at the other end of the spectrum, the medical record appears to be incomplete in documenting services to high utilizers and to rural patients. Continuing improvement in ambulatory claims is needed if this administrative data resource is to be used for quality and provider performance monitoring. It should be possible to substantially improve on the 82% of claims that agree with records on both date and diagnosis, particularly in hospitals where the rate was found to be 75%.

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of other members of the research team, including Neil Powe, M.D., Andrea Gerstenberger, Sc.D., Andrew Baker, and Cheryl Kaplowitz, Sc.D., as well as the Delmarva Peer Review
Organization and the cooperation of providers who made their medical records available to the study.

References