

# Expert Agreement in Current Procedural Terminology Evaluation and Management Coding

Mitchell S. King, MD; Martin S. Lipsky, MD; Lisa Sharp, PhD

**Background:** Available data suggest that physicians are accurate in approximately 55% of *Current Procedural Terminology (CPT)* evaluation and management (E/M) coding for their services. This accuracy is relative to observers' or auditors' assigned codes for these services, a group that has not been studied for their consistency in application of the *CPT* E/M coding guidelines. The purpose of this study was to determine the level of agreement of certified coding specialists in their application of *CPT* E/M coding guidelines.

**Methods:** Three hundred certified professional coding specialists randomly selected from the active membership of the American Health Information Management Association were sent 6 hypothetical progress notes of office visits along with a demographic survey. The study group assigned *CPT* E/M codes to each of the progress notes and completed the demographic survey.

**Results:** Coding specialists agreed on the *CPT* E/M codes for 57% of these 6 cases. The level of agreement for the individual cases ranged from 50% to 71%. Relative to the most common or consensus code, undercoding of established patients occurred more commonly than overcoding. In contrast, for new patient progress notes, overcoding relative to the consensus code was more common than undercoding.

**Conclusions:** There is substantial disagreement among coding specialists in application of the *CPT* E/M coding guidelines. The results of this study are similar to results of prior studies assessing physician coding accuracy, suggesting that the *CPT* coding guidelines are too complex and subjective to be applied consistently by coding specialists or physicians.

*Arch Intern Med.* 2002;162:316-320

**D**URING THE past decade, the Health Care Financing Administration (HCFA) has revised the *Current Procedural Terminology (CPT)* coding guidelines in an effort to clarify the work of physicians. Prior to 1992, fee schedules for physician's services were determined by a customary and reasonable charge method.<sup>1</sup> In 1992, this fee schedule was changed and replaced by a system based on relative-value units and conversion factors. To implement this system, a new *CPT* coding system was used, and in 1995, HCFA developed guidelines for use of this new *CPT* coding system. Use of *CPT* evaluation and management (E/M) codes for a patient visit requires determining the level of history taking, physical examination, and medical decision making for the patient and matching the combination of these 3 elements to the proper *CPT* E/M code. These guidelines provide physicians and insurance carriers a format for determining the proper coding level based on medical record documen-

tation. Continued efforts to refine and standardize the guidelines across specialties led to development of new guidelines in 1997 and plans for additional guidelines in 1999, which are still in development.

In today's climate of health care regulation, the accuracy of how physicians use *CPT* coding to define their E/M services is receiving more attention. Since coding is connected to reimbursement, there is concern that financial incentives might lead to coding inaccuracies. However, inaccuracies in coding might also stem from the complexity of the revised coding systems rather than a financial motivation to overcode.

Current data suggest that physicians code improperly, with conflicting data on the net economic impact of this inaccuracy. Information from HCFA and the American Academy of Family Physicians indicates that family practitioners undercode for services, resulting in a loss of potential revenues.<sup>1</sup> Conversely, the Office of the Inspector General<sup>2</sup> recently issued a release citing \$20 billion of Medicare

From the Department of Family Medicine, Northwestern University Medical School, Chicago, Ill.

## METHODS

The study group consisted of 300 certified coding specialists—physician based selected randomly by the American Health Information Management Association (AHIMA) from active members. The AHIMA is 1 of 2 major professional organizations that provide education and certification programs in medical coding. Coding specialists with the certified coding specialists—physician based status were chosen because this certification indicates training and certified competency through testing in physician office–based *CPT* and *International Classification of Diseases* coding. The membership of AHIMA was chosen because they represent a heterogeneous group, including coding specialists from urban, suburban, and rural settings, as well as from different practice models. In addition, AHIMA endorsed the study and provided a mailing list of the 300 randomly selected active members.

Six cases presented as hypothetical progress notes were developed representing different levels of service as well as new and established patient visits. The following 6 problems were chosen for these progress notes: pneumonia, leg cramps/hypertension, deep vein thrombosis (follow-up), exercise-induced asthma, gastroenteritis, and sinusitis/hypertension. These were selected because they represent common problems encountered by family physicians. A sample note is presented in **Table 1** (copies of the other notes available from the authors on request). The patient cases were labeled as “new” or “established,” and only the appropriate *CPT* codes were provided as choices for selection. For example, codes 99201 through 99205 were provided for cases of new patients and codes 99211 through 99215 were provided for cases of established patients.

These cases were then peer-reviewed by family physician faculty at Northwestern University Medical School for completeness and to assess the authenticity in representing actual patient cases.

In addition, a brief survey was developed with demographic and practice characteristics that might be associ-

ated with coding ability. Items were generated using information derived from the literature and expert opinion. For example, practice location was included in the survey since a prior study indicated that practice location influences physician coding.<sup>8</sup> The survey instrument was piloted among coding specialists from AHIMA for content validity and reliability. Feedback from the coders was then incorporated into a final survey instrument.

The survey instrument and cases were mailed, with a self-addressed return envelope and cover letter to the study participants. The cover letter briefly described the project and contained the endorsement of AHIMA. Because of the potential sensitive nature of coding errors, complete anonymity was ensured. Instructions were provided to complete the survey and to code the office visit cases with a *CPT* E/M code based on the documentation found in the sample progress notes, using the 1997 *CPT* E/M coding guidelines. Participants were allowed to use whatever resources they might typically use in their own practice (eg, books, articles) to code the sample notes. An incentive of \$25 dollars was provided for individuals who completed the survey. After 1 month, nonresponders received a second mailing. Two additional mailings were sent to nonresponders.

The “correct” or consensus *CPT* E/M code was defined as the coding level most commonly agreed on for each case. Coding accuracy was defined as the number of cases coded correctly, or in agreement with the consensus codes for the 6 cases. To compare the coding specialist’s responses on the new cases vs the established cases, a frequency count of the cases coded correctly, overcoded, and undercoded was completed across the 3 new cases and across the 3 established cases. Individual performances were evaluated by summing the number of the 6 cases coded correctly.

Descriptive statistics were used to summarize the sample characteristics. An analysis of variance was used to compare groups when appropriate. Scoring on the new cases vs the established cases were analyzed using the nonparametric Wilcoxon matched pairs test.

overpayments with 29% owing to improper coding for physician’s services. A recent study using trained observers and current *CPT* guidelines found that physicians agreed with the observers’ codes for established patients 55% of the time.<sup>3</sup> Errors were almost equally divided between undercoding and overcoding. Kikano and colleagues<sup>4</sup> noted similar results for established patient visits but found that for new patient visits, physicians tended to overcode. Using medical record auditors, Zuber and associates<sup>5</sup> found a higher level of undercoding for established patient visits than seen in prior studies. However, in this study, the 3 auditors (physician faculty, resident, and professional coder) agreed with each other only 31% (1995 guidelines) and 44% (1998 guidelines) of the time, similar to the interrater reliability findings in the study by Kikano et al.<sup>4</sup>

These studies suggest that despite revisions in the current coding system, physicians continue to have trouble using the *CPT* E/M coding guidelines correctly. One explanation for inaccurate coding may be a system that is

too complex and subjective to be applied uniformly.<sup>6,7</sup> Ultimately, a physician’s coding accuracy is judged by experts who audit physician medical charts and examine if the coding level reflects the documented services provided. This assumes that the experts can apply these codes uniformly. However, despite the financial and legal implications of the assumption that the coding system can be applied consistently by coding specialists, there is little research examining the agreement among expert coders in their interpretation of HCFA guidelines.

In this study, we examined the consistency and variability of the current *CPT* coding guidelines when applied by certified coding specialists to medical records for outpatient visits. In addition, we sought to see if characteristics such as years experience in coding, time per week spent coding, number of records coded per week, and type or location of practice are associated with coding accuracy. The results might help to define the complexities of the coding guidelines as well as assist in determining a natural background error rate for coding. This

**Table 1. Coding Survey Cases\***

The following 6 progress notes represent hypothetical patients seen in a practice. Using the 1997 CPT E/M coding guidelines, please circle the level for the individual coding elements (ie, history, examination, medical decision making) for each visit as well as the final CPT E/M level of service you believe this visit represents.

(circle)	Elements (circle)		
	History	Examination	Medical Decision
99211	History	Examination	Medical Decision
99212	PF	PF	S
99213	EPF	EPF	LC
99214	D	D	MC
99215	C	C	HC

Established patient

Chief complaint: Fever, fatigue

A 45-year-old previously healthy man presents with a 3-day history of feeling feverish and fatigued. He notes fever to 39°C and fatigue to the point of not being able to perform his normal exercise and home chores. He has also noted a slight cough worsened with activity. He denies pain, shortness of breath, upper respiratory congestion, sore throat, ear pain, emesis, diarrhea, dysuria, or rash. He is a 35 pack-years smoker.

Temperature, 38°C

Heart rate, 100

Respiratory rate, 24

Blood pressure, 130/80 mm Hg

Alert in no acute distress

HEENT: tympanic membranes and throat clear, nose without significant congestion or rhinorrhea

Neck: supple without lymphadenopathy

Lungs: left lower lobe with bronchial breath sounds, egophony changes and dullness to percussion, no wheezing

Heart: regular tachycardia without murmurs

Abdomen: bowel sounds present, no masses, organomegaly, costovertebral angle or other tenderness

Chest x-ray with left lower lobe pneumonitis

Pulse oximetry 96% on room air

Assessment: Left lower lobe pneumonia

Plan:

Clarithromycin, 500 mg, twice daily for 10 days

Follow-up in 7 days or as required by worsening symptoms

Will need follow-up chest x-ray to show resolution

\*CPT E/M indicates *Current Procedural Terminology* evaluation and management; PF, problem focused; S, straightforward; EPF, expanded PF; LC, low complexity; D, detailed; MC, moderate complexity; C, comprehensive; HC, high complexity; and HEENT, head, eyes, ears, nose, and throat.

**Table 2. Characteristics of 136 Coding Specialists**

Characteristic	Outcome
Coding experience, mean (SD)	
Years coding experience	10.9 (7.1)
Years experience coding in physician's office	8.3 (6.5)
Hours per week coding	24.9 (14.6)
No. of records coded per week	278 (324)
Type of practice, No. (%)	
Primary care	19 (15)
Specialty	54 (42)
Both	56 (43)
Practice location, No. (%)	
Urban	46 (35)
Suburban	38 (29)
Rural	21 (16)
Mixed/other	25 (18)

may help distinguish between fraudulent billing practices vs the difficulty of applying a complex system with perfect accuracy.

## RESULTS

Of the 300 mailing labels provided, 294 had complete information needed for mailing. Of the 294 surveys sent, 2 were returned as undeliverable, leaving a study group

of 292. A total of 136 of 292 eligible for study returned the survey for a response rate of 46%. Thirty-six responders gave incomplete demographic information; however, coding was completed for the cases. In addition, 5 individuals completed coding for all but 1 of the cases. The results of these individuals are included in the analysis of CPT coding accuracy.

**Table 2** summarizes the characteristics of the study group. As given in Table 2, the group averaged 10.9 years of coding experience, with an average of 8.3 years of experience coding in physicians' offices. The mean number of hours per week spent coding was 24.9 hours. The average number of records coded per week was 278. Fifteen percent of the coders coded only for primary care physicians, 42% only for specialist physicians, and 43% for primary care and specialist physicians. All the coders were certified as certified coding specialists—physician based, and 48.5% had an additional coding certification status. Thirty-five percent of coders were located in urban practices, 29% in suburban practice, and 16% in rural practices.

Coding results of the 6 cases are given in **Table 3**. The agreement among the coders in assigning CPT codes ranged from 50% to 71% across the cases. The level of overall agreement for all of the cases was 58.7%. The frequency of overcoding, undercoding, and correct coding is presented in **Table 4**. New patient progress notes were overcoded in 33% of cases, which is 4 times the rate of

**Table 3. Coding Specialist *Current Procedural Terminology* (CPT) Evaluation and Management Coding of 6 Hypothetical Cases\***

	CPT Coding Level, No. (%)				
	99211	99212	99213	99214	99215
<b>Established Patients</b>					
Case 1: pneumonia	0	7 (5)	47 (35)	78 (59)	1 (1)
Case 2: cramps, HTN	2 (2)	25 (19)	65 (50)	39 (30)	0
Case 3: DVT follow-up	16 (12)	89 (67)	27 (20)	0	0
<b>New Patients</b>	<b>99201</b>	<b>99202</b>	<b>99203</b>	<b>99204</b>	<b>99205</b>
Case 4: asthma	17 (13)	66 (50)	32 (24)	17 (13)	1 (1)
Case 5: gastroenteritis	94 (71)	32 (24)	6 (5)	0	0
Case 6: sinusitis/HTN	17 (13)	74 (56)	35 (27)	6 (5)	0

\*HTN indicates hypertension; DVT, deep vein thrombosis.

undercoding for new patients and twice the rate of overcoding of established patients. Established patient progress notes were undercoded in 25% of cases, which represents 3 times the rate of undercoding for new patient cases. Thus, undercoding occurred significantly more often for established patients, and overcoding occurred significantly more often for new patients ( $P = .001$ ).

The coders' overall score relative to the consensus coding response is as follows:

No. of Cases Coded Correctly by Coding Specialists	Coders, No. (%)
6	9 (6.8)
5	26 (19.5)
4	39 (29.3)
3	21 (15.8)
2	27 (20.3)
1	9 (6.8)
0	2 (1.5)

Seven percent of the coders agreed with the consensus code for all of the cases, and 26% agreed for 5 or more of the 6 cases. Twenty-eight percent of the certified coders were in agreement with the consensus code for less than 50% of the cases; however, 97.7% and 92.5% of responses were within 1 coding level of the consensus response for established patients and new patients, respectively.

Coding accuracy (ie, number of cases coded correctly) was not significantly correlated with years of coding experience, years coding in physicians' offices, practice type, or location. Coding accuracy was correlated  $-0.31$  ( $P < .001$ ) with hours spent coding and  $-0.36$  ( $P < .001$ ) with number of records coded per week.

### COMMENT

The results of this study suggest that certified coding specialists do not agree on codes using current CPT guidelines. This is a particularly troublesome finding given that the coding specialists involved in this study were all certified by at least 1 professional coding organization and 25% of the coders held certificates from both professional coding organizations.

The experts' codes for established patients were in agreement for 58% of the cases coded, findings similar to a recent study in which physicians' codes for established visits agreed with that of a trained observer 55% of the time.<sup>3</sup> The level of agreement among the coders for new patients

**Table 4. Undercoding and Overcoding of Cases by Coding Specialists**

Patients	Undercoded, %	Agreement, %	Overcoded, %
Established	24.5*	58.6	16.9*
New	8.5*	58.9	32.5*

\* $P = .001$ .

was also 58%, slightly better than the 48% found for physicians.<sup>4</sup> Taken together, these data would suggest that physicians' coding accuracy is not much different from that of trained certified coding specialists.

In addition, the patterns of errors in our study with expert coders were similar to prior physician studies.<sup>3-5,9</sup> In our study and others, undercoding is more common in cases of established patients whereas overcoding is more common with new patients. One reason for this discrepancy could be a tendency to apply the same guidelines to all patients, not recognizing or applying the different criteria for new patients. Coding criteria are stricter for new patients, requiring more documentation to establish the same service level. In addition, physicians and coding specialists alike may recognize that caring for new patients requires more effort and that there is more uncertainty in providing this care than for established patients. Thus, physicians and coders may feel that new patients are more difficult and coding levels may reflect this rationale.

Although one might predict that experience would improve coding accuracy, this study found no such association. In addition, no associations between coding accuracy and type of practice or practice location were found. A negative correlation was found between number of hours spent coding and numbers of records coded per week. This suggests that excessive time spent coding and higher volumes of coding may actually compromise accuracy. Another explanation may be that individuals who spend many hours per week coding might tend to work for larger practice groups. In today's climate of potential audits, these larger organizations could create a conservative coding atmosphere that promotes a tendency toward undercoding.

Although coding errors might conceivably relate to financial incentives or potential legal penalties, the for-

mat of the study was designed to test the coding specialists' accuracy in coding using hypothetical cases. This design removes any financial or legal incentives for incorrect coding. All coding specialists coded from the same type-written progress notes, thus removing the discrepancies from attempting to interpret handwritten progress notes or apply the guidelines to different cases of the same coding level. Despite removing these potential sources of coding inaccuracy, the error rate was still high, with 44% of coding specialists agreeing with the consensus response on 3 or fewer cases, and 8% agreeing with the consensus code on 1 or none of the cases. However, only 3% of established patient codes and 8% of new patient codes were more than 1 coding level different from the consensus code. Thus, although there seems to be a high background error rate for CPT coding among coding specialists, most errors are within 1 level of the correct code. This finding is consistent with findings from the physicians coding study by Kikano et al<sup>4</sup> who found that physicians' codes differed from reviewers' codes by more than 1 level in fewer than 4% of cases. Unless this intrinsic coding error rate is accounted for, identification of fraudulent coding practices would be extremely difficult.

From our results, it seems that the error rate with CPT coding is substantial for coding specialists as well as physicians. This would suggest that the guidelines themselves are overly complex and open to subjective interpretation which then creates a high inherent error rate. Having separate sets of guidelines for new and established patients may be a contributory factor. One possible solution to minimizing the error rate with CPT coding would be to standardize the coding criteria into 1 set of guidelines for all patients. In addition, decreasing the number of potential codes for each office visit as well as the number of steps required to arrive at a code would limit the potential for error and subjective interpretations. Another proposed solution<sup>6</sup> involves using time and new vs established patient status as the deciding factors in arriving at the level of ser-

vice provided. Finally, given the complexity of the current CPT guidelines, another potential solution is to accept an inherent error rate. Clearly, further study of the CPT coding guidelines is warranted.

Accepted for publication May 8, 2001.

This study was funded by a research grant from the American Academy of Family Physicians Foundation, Leadwood, Kan.

We thank the American Academy of Family Physicians Foundation for their generous grant support and the American Health Information Management Association, Chicago, Ill, for their assistance.

Corresponding author and reprints: Mitchell S. King, MD, Glenbrook Family Care Center, 2050 Pfingsten Rd, Room 200, Glenview, IL 60025 (e-mail: m-king1@nwu.edu).

## REFERENCES

1. Sgammato J. HCFA answers questions about its new documentation guidelines. *Fam Pract Manag.* 1995;2:60-67.
2. Martin S. OIG: \$20 billion in "improper" Medicare payments. *American Medical News.* May 11, 1998:11-13.
3. Chao J, Gillanders WG, Flocke SA, Goodwin MA, Kikano GE, Stange KC. Billing for physician services: a comparison of actual billing with CPT codes assigned by direct observation. *J Fam Pract.* 1998;47:28-32.
4. Kikano GE, Goodwin MA, Stange KC. Evaluation and management services: a comparison of medical record documentation with actual billing in a community family practice. *Arch Fam Med.* 2000;9:68-71.
5. Zuber TJ, Rhody CE, Muday TA, et al. Variability in code selection using the 1995 and 1998 HCFA documentation guidelines for office services. *J Fam Pract.* 2000; 49:642-645.
6. Lasker RD, Marquis MS. The intensity of physicians' work in patient visits: implications for coding of patient evaluation and management services. *N Engl J Med.* 1999;341:337-341.
7. Iezzoni LI. The demand for documentation for medicare payment. *N Engl J Med.* 1999;341:365-367.
8. Purvis JR, Horner RD. Billing practices of North Carolina family physicians. *J Fam Pract.* 1991;32:487-491.
9. Horner RD, Paris JA, Purvis JR, Lawler FH. Accuracy of patient encounter and billing information in ambulatory care. *J Fam Pract.* 1991;33:593-598.