

We try to publish authors' responses in the same edition with readers' comments. Time constraints might prevent this in some cases. The problem is compounded in a bimonthly journal where continuity of comment and redress are difficult to achieve. When the redress appears 2 months after the comment, 4 months will have passed since the article was published. Therefore, we would suggest to our readers that their correspondence about published papers be submitted as soon as possible after the article appears.

### Maternal Smoking and Congenital Birth Defects

*To the Editor:* In their retrospective study of maternal smoking and the risk of congenital birth defects, Woods and Raju<sup>1</sup> stated that they "used the Bonferroni adjustment (significance at  $P < .01$  for each outcome, for an overall significance of  $P = .05$ ) to reduce the possibility of chance playing a significant role." Setting the statistical significance level for individual comparisons at .01 adjusts for exactly 5 comparisons by the Bonferroni method,<sup>2</sup> which corrects the statistical threshold for multiple comparisons by simply dividing the significance level (alpha) by the number of comparisons made ( $.05/5 = .01$ ). Woods and Raju, however, compared the occurrence of congenital birth defects in the children of smokers and nonsmokers across 22 different categories of birth defect. Hence, use of the Bonferroni adjustment method would require that the significance level for individual comparisons be set at .002 ( $.05/22 = .002$ ) to maintain an overall significance level of .05.

Based on the significance threshold of  $P < .002$ , maternal smoking was not associated with a statistically significant difference in any of the 22 categories of congenital birth defect in their study, although Woods and Raju reported that congenital anomalies of the cardiovascular system were significantly more frequent in children of smokers ( $P < .01$ ). My reanalysis of their data indicates that, after applying the Bonferroni adjustment, the increased occurrence of congenital anomalies of the cardiovascular system in children of smokers was slightly above the threshold of statistical significance ( $P = .0025$ ) by the chi-square test. The observed difference might possibly be clinically significant, but not quite statistically significant.

Some experts have argued that the Bonferroni adjustment is overly conservative, setting the threshold of statistical significance too high. In addition, the custom of requiring a  $P$  value less than .05 for individual comparisons has also been challenged as being arbitrary and too rigid. If the authors wish to claim an overall significance level of .05 when making multiple comparisons, however,

it is important to apply the appropriate statistical correction, seeking input from a statistician if required.

Thomas M. Morgan, MD  
Yale University School of Medicine  
New Haven, Conn

### References

1. Woods SE, Raju U. Maternal smoking and the risk of congenital birth defects: a cohort study. *J Am Board Fam Pract* 2001;14:330-4.
2. Fisher LD, van Belle G. *Biostatistics: a methodology for the health sciences*. New York: John Wiley, 1993:611.

The above letter was referred to the authors of the article in question, who offer the following reply.

*To the Editor:* In his letter, Dr. Morgan questions the way the Bonferroni adjustment was used in the article "Maternal Smoking and the Risk of Congenital Birth Defects." He is correct in pointing out that dividing the alpha (.05) by 22 (the number of comparisons) would set the new significance level at .002 to maintain an overall significance level of .05. He is also correct in pointing out that most statisticians feel that the Bonferroni adjustment is too conservative, setting the level of significance too high.

Using the raw data in Table 3, Dr. Morgan calculated a  $P$  value of .0025 for the cardiovascular congenital defects using a chi-square test. Although the raw data are presented in Table 3, the reported relative risks and  $P$  values are adjusted for the confounders of age, race, and diabetes. This analysis is multivariate. The resulting  $P$  value from the regression analysis was .0009, smaller than the  $P$  value from the univariate analysis as reported by Dr. Morgan. This category is still significant even if you set the adjusted alpha at .002.

It is my opinion, as well as the opinion of the primary statistician that I have used in the past, that the Bonferroni adjustment is too conservative. Although I believe that adjusting the alpha for multiple comparisons is appropriate, I also believe that setting the alpha below .01 is setting the threshold of significance too high. The  $P$  value for our study was set at .01 to reduce the risk of chance with multiple comparisons without making statistical significance prohibitively difficult.

The results are statistically significant for the cardiovascular congenital defects regardless of which  $P$  value is used. The authors concluded, however, similar to what other authors have concluded, that based on this study and all of the epidemiologic evidence on this subject, that gestational smoking does not cause any significant increase in birth defects. It is interesting that Shiono and colleagues,<sup>1</sup> in the prospective Kaiser Permanente Birth

Defects Study, found a statistically significant protective effect of smoking and ventricular septal defect.

Scott E. Woods, MD, MPH, MED  
Bethesda Family Practice Program  
Cincinnati, Ohio

## Reference

1. Shiono PH, Klebanoff MA, Berendes HW. Congenital malformations and maternal smoking during pregnancy. *Teratology* 1986;34:65-71.

## Evaluation and Management Coding by Family Physicians

*To the Editor:* As faculty members who are responsible for teaching family practice residents how to code evaluation and management visits appropriately, we read with interest the article by King and colleagues entitled "Accuracy of CPT Evaluation and Management Coding by Family Physicians" (King MS, Sharp L, Lipsky MS. *J Am Board Fam Pract* 2001;14:184-92). The authors showed that for new patients, family physicians agreed with the experts only 17% of the time, the predominant error being overcoding. They concluded, "the error rate for physician CPT coding is substantial."

Because the documentation guidelines used for evaluating the six sample progress notes was not mentioned, we assume either the 1995 or the 1997 documentation guidelines could be utilized. If such were the case, by using the 1995 documentation guidelines, we would have to disagree with 3 of the 6 interpretations by the professional coders.

Case 2 represented an established patient with stable hypertension and a new complaint of leg cramps. The history is detailed (D) (3 HPI elements, 2 to 9 ROS, and SH [nonsmoker]). The examination is detailed (D) (5 elements evaluated). The decision making was moderately complex (MC) in that according to the diagnosis and management options score, 1 point is awarded for "established, previous diagnosis" (hypertension), and 3 points for "new problem, no additional workup planned" (leg cramps). Risk level was moderate because of a "new problem without a diagnosis" (leg cramps) and prescribed medications (atenolol and hydrochlorothiazide). The combination of D/D/MC corresponds to a 99214 level visit, not 99213 as reported by the professional coders.

Case 5 represented a new patient with diarrhea. The history is expanded problem focused (EPF) (1 to 3 HPI elements, 1 ROS, and PMH [no medications]). The examination is also problem focused (EPF) (4 elements evaluated). The decision making was straightforward (S) in that 1 point is awarded under diagnosis and management options with minimal risk level ("self-limited, minor problem"). The combination of EPF/EPF/S corresponds to an evaluation and management level of 99202, not 99201 as reported by the professional coders.

Case 6 represented a new patient with a sinus infection. The history is detailed (D) (3 HPI elements, 9 ROS ["ROS otherwise negative"], and PMH/SH). The examination is detailed (D) (5 elements evaluated). The deci-

sion making was moderately complex (MC). Under diagnosis and management options, 1 point is awarded for "previous, established diagnosis" (hypertension), and 3 points for a "new problem, no additional workup planned" (sinusitis). Risk level was moderate because medication was prescribed (captopril and amoxicillin). The combination of D/D/MC corresponds to an evaluation and management level of 99203, not 99202 as reported by the professional coders. If decision making were rated low complexity by considering sinusitis a "self-limited, minor problem," the result would still be a 99203 level visit.

This comparison serves to highlight several points. First, the Illinois physicians did not overcode new patients as badly as reported, although we would agree there is generally a tendency to overcode new patients compared with established patients. Second, the current procedural terminology (CPT) level can differ depending on whether the 1995 or 1997 documentation guidelines are used. Third, CPT coding is open to interpretation. Last, because it is possible to arrive at different conclusions when coding a particular note, the current system of CPT coding is seriously flawed. We can only hope the new guidelines currently being evaluated will remedy these flaws.

Roland Larrabee, MD  
Walter L. Aument Family Health Center  
Quarryville, Pa  
Lancaster General Hospital  
Lancaster, Pa  
Sherry Hess, MBA, RHIA, CCS-P  
Lancaster General Hospital  
Lancaster, Pa

## Association and Characteristics for Screening Mammography

*To the Editor:* While reviewing Table 3, "Univariate Associations of Visit Characteristics and Physician Recommendation for Screening Mammography," in the article "What Influences Mammography Recommendations?" (Nutting PA, Baier M, Werner JJ, et al. 2001;14:352-61.), one disturbing question leaped out at me. Are we helping or hindering our patients' access to preventive services?

According to this study, of all factors contributing to whether a woman is referred for mammography, the most substantial was a patient's request for a mammogram. Prompting from a patient had far greater effect on mammography referrals than any other physician or patient characteristic, save the patient with a personal or family experience with breast cancer (which might be another form of patient request).

Given a traditional medical culture in which the "physician proposes and patient disposes," there might be many more patients who desire screening mammography but are not referred because the physician never proposes one. This group (1 - the squeaky wheels) likely has the same breast cancer risk as those who (gently or not) speak up and remind us to order their mammograms.

Because of respondent bias (the physicians who are most likely to refer patients for mammograms are most likely to volunteer for a study of physician referral rates for mammography), the discrepancy between actual mammography referral rates and guideline-based goal rates is likely much larger than that detected by the competing demands model study. The preventive care services gulf widens when the differences between referrals for services and completion of the services are considered (those who keep tickler files well know the gap between those who are referred and those who follow through).

A possible interpretation of this finding is that our patients might know or care more about preventive care than we do, though I doubt it is so. It is likely that our organizational structures, institutional inertia, paternalistic heritage, and nonsystem of health care reimbursement impede our ability to initiate and follow through with preventive care services.

Peter Teichman, MD, MPA  
West Virginia University  
Harpers Ferry, WV

The above letter was referred to the authors of the article in question, who offer the following reply.

*To the Editor:* Dr. Teichman questions whether we are helping or hindering our patients' access to preventive services. While we did not directly address this issue in the study, his thoughtful question deserves consideration across a broad range of preventive services. In reporting our study of screening mammography, we specifically termed the event a physician recommendation for mammography, rather than a referral, so we could accommodate the considerable regional variation in requiring a formal physician referral among mammography centers and health plans. Our study captured the determinants of physician recommendation for screening mammography during patient visits and did not include, for example, occasions in which patients either go directly to the mammography center or telephone the practice to obtain a required preauthorization. We made this decision consciously so we could compare the relative importance of physician characteristics with the specific visit characteristics.

While we do not believe that failure of physicians to recommend a mammogram necessarily hinders access for patients who want to be screened, we do agree (based on our data and that of others) that we often do not do enough to encourage screening mammography. Our study examines and shows that other patient needs for service during the visit apparently compete for attention on the patient-physician visit agenda. The importance of Dr. Teichman's question, however, might vary across a range of preventive services that do require specific physician action to enable patient access to the procedure.

We also agree with Dr. Teichman's assertion that a number of system variables hinder appropriate provision of clinical preventive services. Recent work by other investigators has begun to unravel the black box of primary care practice to better understand how to improve preventive services.<sup>1,2</sup> Our study underscores the importance of competing demands and opportunities among the variety of services provided in primary care and emphasizes the important observation that visit characteristics are more predictive of preventive services than are physician knowledge and attitudes.

We hope that our findings will stimulate further attention to the structures and processes of primary care practice, and reduce the tendency to blame the primary care physician entirely for failing to provide whatever service may be the focus of any particular study.

Paul A. Nutting, MD, MSPH  
Colleen Conry, MD  
James J. Werner, MS

University of Colorado Health Sciences Center,  
Denver

Monika Baier, MS

Gary Cutter, PhD

AMC Cancer Center, Denver

Linda Stewart, MD

Family Medicine Center of Baton Rouge,  
Baton Rouge, La

## References

1. Stange KC, Zyzanski SJ, Jaen CR, et al. Illuminating the 'black box'. A description of 4,454 patient visits to 138 family physicians. *J Fam Pract* 1998;46:377-89.
2. Crabtree BF, Miller WL, Stange KC. Understanding practice from the group up. *J Fam Pract* 2001;50:881-7.