In Pursuit of Compatible Coding Audit Benchmarks

Save to myBoK

By Mary H. Stanfill, MBI, RHIA, CCS, CCS-P, FAHIMA

People like benchmarkS. They like to know how they are doing in comparison to someone else. In health information management (HIM), a common benchmark is the de facto standard of 95 percent accuracy rate for medical coding. HIM professionals regularly measure coding accuracy and compare accuracy rates for medical coders in facilities within a health system. External benchmarking of coding accuracy, however, is more difficult. This is partially due to data privacy requirements as well as concerns for potential exposure to non-compliance in healthcare claims data. But of equal concern is how coding accuracy rates are calculated and the comparability of the data from one healthcare system to another.

It has been suggested that coding accuracy is subjective and must be measured within the context of a given facility.[1] This article focuses on this challenge and compares data derived from various methods for calculating coding accuracy rates.

Comparison of Coding Accuracy Rates

The most common methods of calculating coding audit results are either per code or per record, and the statistical equation applied may include weighting of audit elements. Results can vary widely and may or may not be comparable depending on which method is used.

Code-over-code accuracy considers the number of codes that are originally assigned correctly divided by the total correct codes. Historically, a non-weighted code-over-code approach was favored because code variances are counted equally with no judgement on the importance of a particular code variance. But weighted scoring mechanisms are becoming more common to incorporate the nuances of well-established reimbursement methodologies and compliance requirements.

Record-over-record accuracy considers the number of records with correct coding divided by the total number of cases reviewed. Sometimes referred to as the "all right/all wrong method,"[2] the non-weighted record-over-record calculation is useful for binary audit elements (such as DRG accuracy) but makes it challenging to achieve a high score when multiple audit elements are considered on each record. In contrast, weighted record-over-record calculations can account for multiple audit elements and present a more achievable overall score.

To illustrate the differences, the author of this article applied each method to the same random sample of 25 inpatient cases. The 25 cases included a total of 475 ICD-10-CM diagnosis codes and 68 ICD-10-PCS procedure codes. A comparison of the accuracy rates derived from each methodology is presented in Table 1 below.

Table 1: Comparison of Accuracy Calculations

Calculation Method	Diagnosis Accuracy	Procedure Accuracy
Non-weighted code-over-code		
• Numerator = Total number of original codes assigned correctly	**92.42%** (=439 correct original codes/475 total codes)	**91.18%** (=62 correct original codes/68 total codes)

• Denominator = Count of correct codes		
Weighted code-over-code		
• Principal diagnosis codes and MCC/CC codes are given twice the weight of additional diagnosis codes • Otherwise, the numerator and denominator are calculated as above	**87.68%** • (76 original principal/ MCC/CC codes correct*2) + 346 original additional Dx codes correct = 498 • (93 principal/MCC/CC codes correct*2) + 382 correct additional Dx codes = 568 • 498/568 = 87.68%	**92.59%** • (13 original principal procedure codes correct*2) + 49 original additional procedure codes correct = 75 • (13 principal procedure codes correct*2) + 55 correct additional codes = 81 • 75/81 = 92.59%
Non-weighted record-over-record		
• Numerator = Total number of cases correct on all (or specified) audit elements as originally presented • Denominator = Number of cases	Overall case rate = **28%** (=7 cases with no errors on any audit element/25 cases) DRG accuracy rate = **88%** (=22 correct original DRGs/25 cases)	
Weighted record-over-record		
• Highest weight: principal diagnosis, secondary diagnoses, POA, principal procedures, or discharge disposition codes that change the DRG • Modest weight: secondary diagnosis codes that change the SOI, ROM, or HCC • Lowest weight: Non-MCC/CC secondary diagnosis, secondary procedure codes, POA, and discharge disposition changes that do not impact DRGs	Overall record accuracy = **91.6%** (=2,290 points scored based on original codes/2,500 possible total points)	

• No weight: educational code notes		

The variety of bolded accuracy rates in Table 1 demonstrate how much accuracy rates vary, depending on how the rate is calculated. For this particular sample of 25 cases, the diagnosis coding appears to be slightly stronger than procedure coding, according to the non-weighted code-over-code rate (92.42 percent). However, the three DRG changes were all due to diagnosis code changes and several of the secondary diagnosis codes changed were MCC/CC (major complication or comorbidity/complication or comorbidity) designated codes. In fact, the majority of the findings were diagnosis code changes, additions or deletions, a fact that is better reflected in the weighted code-over-code rate (87.68 percent).

Analysis of procedure coding detailed results is also more consistent with the weighted code-over-code rate (91.18 percent compared to 92.59 percent). The six procedure code errors were primarily the addition of reportable procedures; appropriate, but not impactful on reimbursement or quality. There were far fewer procedure codes, making the denominator smaller, thus increasing the statistical impact of the changes in a non-weighted code-over-code rate. The non-weighted record-over-record accuracy rate (28 percent) illustrates the difficulty in achieving accuracy when multiple audit elements must be met on each record. This calculation is very useful, however, to measure whether the single DRG audit element is correct on a case—or not (88 percent). The weighted record-over-record calculation is more complex but presents a more thorough overall analysis of the 25 cases, taking into consideration, for example, two cases with risk of mortality (ROM) changes and one case with an inaccurate discharge disposition (none of which impacted the DRGs).

Non-weighted accuracy rates are particularly useful to measure performance on specific codes or groups of codes in a code set. They are useful for both internal and external benchmarking because the calculation is based on objective code counts. In contrast, weighted accuracy rates allow for more sophisticated benchmarking that incorporates defined priorities, which is extremely useful as follow up action plans depend on the types and severity of errors. Weighted accuracy rates are presently limited for external benchmarking due to the lack of standardization in the weighting used in the equation.

Recommendations for Setting Benchmarks

When determining the methods for calculating coding accuracy and analyzing and interpreting benchmarks, HIM professionals should be mindful of the following:

- Collect data at a detailed level for each audit element to ensure source data is sufficient to calculate the results in multiple ways
- Clearly define the numerator and denominator for an accuracy rate and collect data consistently
- Use weighted and/or non-weighted methods intentionally and consistently
- Interpret accuracy results within context, considering how the rate was calculated
- Voluntarily collaborate with other healthcare systems to advance development of external benchmarks

A literature review of benchmarking in healthcare highlights how benchmarking approaches have evolved in the healthcare industry. The authors of a Healthcare Policy article observe "…this evolution produced numerous definitions, whose common theme is continuous measurement of one's own performance and comparison with best-performers to learn about the latest work methods and practices in other organizations."[3] HIM professionals should continue to pursue useful coding accuracy measurements internally but also seek and voluntarily contribute to external benchmark data for continuous improvement in the medical coding function.

Notes

1. Chapman, Susan. "Best Practices in Coding Audits." *For the Record* 30, no. 1 (January 2018): 10.
2. Humbert, Sarah. "How to Choose the Right Coding Audit Method." *Journal of AHIMA* 89, no. 3 (March 2018): 18-19.
3. Ettorchi-Tardy Amina, Marie Levif, and Philippe Michel. "Benchmarking: A Method for Continuous Quality Improvement in Health." *Healthcare Policy* 7, no. 4 (May 2012): 101-119.

Mary H. Stanfill (mstanfill@uasisolutions.com) is vice president of consultant services at United Audit Systems, Inc.

Article citation:
Stanfill, Mary H. "In Pursuit of Comparable Coding Audit Benchmarks." *Journal of AHIMA* 90, no. 1 (January 2019): 30-31, 47.