Preparing for ICD-10-CM/PCS Implementation:
Impact on Productivity and Quality

by Mary H. Stanfill, MBI, RHIA, CCS, CCS-P, FAHIMA; Kang Lin Hsieh, MS; Kathleen Beal, MPA, CPHQ, RHIA; and Susan H. Fenton, PhD, RHIA, FAHIMA

Abstract

Coding productivity is expected to drop significantly during the lead-up to and in the initial stages of ICD-10-CM/PCS implementation, now expected to be delayed until October 1, 2015. This study examined the differences in coding productivity between ICD-9-CM and ICD-10-CM/PCS for hospital inpatient cases matched for complexity and severity. Additionally, interrater reliability was calculated to determine the quality of the coding. On average, coding of an inpatient record took 17.71 minutes (69 percent) longer with ICD-10-CM/PCS than with ICD-9-CM. A two-tailed *T*-test for statistical validity for independent samples was significant ($p = .001$). No coder characteristics such as years of experience or educational level were found to be a significant factor in coder productivity. Coders who had received more extensive training were faster than coders who had received only basic training. Though this difference was not statistically significant, it provides a strong indication of significant return on investment for staff training time. Coder interrater reliability was substantial for ICD-9-CM but only moderate for ICD-10-CM/PCS, though some ICD-10-CM/PCS cases had complete interrater (coder) agreement. Time spent coding a case was negatively correlated with interrater reliability (−0.425 for ICD-10-CM and −0.349 for ICD-10-PCS). This finding signals that increased time per case does not necessarily translate to higher quality. Adequate training for coders, as well as guidance regarding time invested per record, is important. Additionally, these findings indicate that previous estimates of initial coder productivity loss with ICD-10-CM/PCS may have been understated.

Key words: ICD-10 implementation, productivity, coding quality

Background

Experienced professional coders have achieved a certain level of productivity in assigning International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes. Conventional wisdom and reports from the implementation of the International Classification of Diseases, Tenth Revision (ICD-10) in other countries indicate that the current coding workforce will not be able meet ICD-9-CM coding productivity standards when coding with International Classification of Diseases, Tenth Revision, Clinical Modification, and International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-CM/PCS). Canada reported a decrease of approximately 50 percent, with ICD-10 productivity never fully recovering to ICD-9 levels.[1] The report from Australia did not specify the productivity decrease, though Australia reported a rebound to ICD-9 coding productivity levels after approximately 12 weeks.[2] Thus an analysis of coder productivity using ICD-10-CM/PCS is needed in order to anticipate staffing needs to successfully make the transition to the new code sets.

Introduction

Productivity loss, or an increased amount of time to code a patient encounter, due to the transition to ICD-10-CM/PCS is expected to reflect a bell curve, with the peak productivity loss surrounding the go-live date. It is presumed that organizations will experience some productivity loss with ICD-9-CM in the weeks leading up to the go-live date as coding professionals spend time and energy on learning the ICD-10-CM/PCS classification systems. When coders first begin actively coding charts in ICD-10-CM/PCS, their production, that is, the throughput of charges on individual encounters, will be slowed significantly. This slowdown is expected to gradually improve as proficiency with ICD-10-CM/PCS improves, eventually leveling out to some "new normal" productivity standard.

If this conventional wisdom is true and coding professionals will not be able to code as fast with ICD-10-CM/PCS, by how much will productivity actually decline? UASI funded this time study, in partnership with the University of Cincinnati and the School of Biomedical Informatics at the University of Texas Health Science Center at Houston (UT-SBMI), to answer this question. The study objectives were to:

- Determine the variance in coder productivity using ICD-9-CM versus ICD-10-CM/PCS coding classifications;
- Calculate the potential initial productivity loss due to the transition to ICD-10-CM/PCS; and
- Evaluate the relationship between ICD-10-CM/PCS coding productivity and quality.

The study hypotheses were as follows:

H_{01} No difference in coding productivity will be found when coders utilize ICD-9-CM and ICD-10-CM/PCS classification systems; OR

H_{A1} Coding productivity using the ICD-10-CM/PCS classification system will be significantly lower than coding productivity using the ICD-9-CM classification system.

H_{02} A reduction in coder productivity will result in an increase in ICD-10-CM/PCS coding quality; OR

H_{A2} A reduction in coder productivity will result in a decrease in ICD-10-CM/PCS coding quality.

The data were collected by UASI, with the coding performed by UASI employees. After the data were cleaned by UASI, the information was turned over to UT-SBMI for data analysis. UASI desired the results for internal planning purposes and agreed to widespread publication of the results for the benefit of the coding industry.

Methods

The study was designed to effectively simulate coding practice and reliably measure the time required to fully code an inpatient health record. The decision was made to have study participants code distinct but similar cases in only one code set and compare the two. The authors determined that it was essential to factor in the portion of time related to reading the health record to identify diagnoses and procedures that warrant coding and reporting. If an inpatient case were coded in both ICD-9-CM and ICD-10-CM/PCS, the time to become familiar with the case would be reflected in the results for whichever code set was used first. Creating two groups of similar inpatient cases and coding one group with ICD-9-CM and the other with ICD-10-CM/PCS ensured that the necessary record analysis time did not skew coder productivity results. This method more closely reflects the actual transition that will occur operationally;

that is, hospital coding staff will code inpatient cases in ICD-9-CM and then switch to coding inpatient cases in ICD-10-CM/PCS, but for the most part, these cases will only be similar, not identical, with duplicate coding of cases occurring in a limited number of instances.

Table 1 describes additional elements affecting coding productivity and how each was addressed in designing the study.

Any analysis of productivity is useful only if the accuracy is validated in some manner to ensure that accuracy is not sacrificed in the quest for speed. The following options for measuring participant accuracy were considered:

- An expert gold standard
- Consensus-based standard
- Calculation of interrater reliability (IRR)

Ultimately, the calculation of an IRR score was determined to be most informative. IRR was calculated for both the ICD-9-CM and ICD-10-CM/PCS coding. In the case of ICD-10-CM/PCS the IRR calculation included the ICD-10 gold standard coders as described in the Data Collection section.

Sample Selection

To measure both of the new code sets, the study focused on inpatient acute care cases, for which both ICD-10-CM and ICD-10-PCS codes will be reported once the transition occurs. The two groups of records coded in the study were carefully selected by University of Cincinnati Health (UC). IRB approval was sought, with this study designated as exempt. The type and complexity of cases were based on the most frequently reported Medicare Severity Diagnosis Related Groups (MS-DRGs), starting from a list of the top 25 MS-DRGs reported in recent months. The record samples were also carefully constructed to represent both medical and surgical cases. See Table 2 for cases from the MS-DRGs included in the study.

UC staff selected cases of these types and complexity that were complete and closed, with all clinical documentation available to the study participants. Two groups of sample cases, consisting of 27 equivalent inpatient cases in each group, were created. Equivalence was accomplished by identifying a case to include in each group for each MS-DRG that had the same length of stay and similar severity of illness in terms of the number of conditions to code and report.

Study Participants

All of the study participants were ICD-9-CM coding experts with many years of experience in coding inpatient health records. The preliminary data presented here are based on the results of the six participants organized into two groups: a basic group and an advanced group. The basic group consisted of three people who had received only basic orientation to coding with ICD-10-CM/PCS, which amounted to approximately 12 hours of training. The advanced group consisted of three people who were all American Health Information Management Association (AHIMA)–approved ICD-10 Trainers, which amounts to approximately 50 hours of training.

Data Collection

Study participants coded the 27 records in group A with ICD-9-CM and the 27 records in group B with ICD-10-CM/PCS using the same encoder tool. Additionally, each ICD-10-CM/PCS record was coded by an "ICD-10 gold standard coder." There were two ICD-10 gold standard coders, one who coded 13 records and one who coded 14 records. Both of these coders were AHIMA employees who had developed and were involved in delivering the AHIMA ICD-10 Train-the-Trainer workshops. ICD-9-CM has been used for so long in the US that it was not felt to be useful to utilize ICD-9-CM gold standard coders. The inpatient cases in each record set were coded no more than once by a study participant. However, each inpatient case was coded multiple times. The time (in total minutes) to code a case, all codes assigned, and narrative notes on any issues related to clinical documentation or code specificity

were recorded for each case. Study participants were not limited to a certain number of diagnoses or procedures to capture per inpatient case, but rather were directed to code all reportable conditions and procedures consistent with Uniform Hospital Discharge Data Set (UHDDS) reporting guidelines. A standardized Excel data collection tool was designed to capture this information from each study participant. The time required to record information in the Excel file was excluded from the coding time.

Once a sufficient number of participants had completed the entire group of cases, individual participant Excel files were reviewed to clean up any data entry errors. Individual de-identified data files were subsequently uploaded to SPSS for statistical analysis.[3] The IRR analysis was conducted using Microsoft Excel.[4]

Results

The average coding time with ICD-9-CM was 25.52 minutes, whereas the average coding time with ICD-10-CM/PCS was 43.23 minutes. This finding means that, on average, coding an inpatient record with ICD-10-CM/PCS took 17.71 minutes (69 percent) longer than coding an inpatient record with ICD-9-CM. A two-tailed *T*-test for statistical validity for independent samples was significant ($p = .001$).

The early data analysis also shows a wide range in the study participants' productivity. Six participants took from 10.6 minutes (31.6 percent) longer to 27.78 minutes (89.8 percent) longer on average when coding in ICD-10-CM/PCS than when coding in ICD-9-CM.

Three participants who had earned the AHIMA-approved ICD-10 Trainer designation took an additional 14.7 minutes (54.4 percent) longer per case on average. Three participants who received only basic ICD-10 training needed an additional 21.0 minutes (87.8 percent) longer on average. One coder with basic training doubled the time required when changing from coding in ICD-9-CM to coding in ICD-10-CM/PCS. This coder was eliminated from the final analysis because this study is applied research and should consider the acceptable range in a production environment. Without this coder, the basic coders needed an additional 17.0 minutes (81.8 percent) longer on average.

Additional findings based on study participant differences related to education and ICD-9-CM coding experience are presented in Table 3.

Nonparametric statistical tests (Kruskal-Wallis and Mann-Whitney U) for the impact of coder characteristics, including coding credential, coder educational level, ICD-9-CM coding experience, hours of ICD-10-CM/PCS training, and experience with ICD-10-CM/PCS coding all failed to reach any level of significance.

As stated previously, IRR was chosen to judge the quality of the coding because it is the most objective measure. The IRR between the coders for ICD-9-CM diagnostic coding was a Kappa value of 0.6783. The IRR for ICD-9-CM Volume 3 (procedure codes) was a Kappa value of 0.6110. The results demonstrate substantial agreement. The calculated IRR for ICD-10-CM was a Kappa value of 0.4857, with an IRR for ICD-10-PCS calculated at a Kappa value of 0.4170. Both of these show moderate agreement. To determine whether IRR improved as more time was spent coding a record, a correlation was run for both the ICD-10-CM results and the ICD-10-PCS results. The time to code a record in ICD-10-CM was found to be negatively correlated with the IRR at −0.425. The time to code a record in ICD-10-PCS was found to be negatively correlated with the IRR at −0.349.

Discussion

These findings indicate that previous estimates of initial coder productivity loss may have been understated. In the absence of any definitive data, many commentators have relied on reported productivity losses from other countries that have made the transition to a version of ICD-10, such as Canada. The prevailing estimate of productivity loss is typically somewhere between 30 and 50 percent. However, preliminary results of this time study overall were much higher, with ICD-10-CM/PCS coding taking 69 percent longer overall and, at best, 54.4 percent longer when performed by the participants with the most training in ICD-10-CM/PCS.

Of particular importance is the strong indication of a significant return on investment for staff training time. Study participants with more than a week of ICD-10-CM/PCS training (that is, those holding the AHIMA-approved ICD-10 Trainer designation) experienced the lowest group average productivity loss at 54.4 percent, while those with only 10 to 12 hours of ICD-10-CM/CPS training experienced an 81.8 percent productivity loss. Though not statistically significant in this limited sample size, the practical significance is considerable for designing effective training.

The wide range of productivity losses experienced by the six participants, ranging from a low of 31.6 percent to a high of 89.8 percent longer on average, is also notable. In addition, comparing individual participants' speed when coding with ICD-9-CM to their speed when coding with ICD-10-CM/PCS seems to indicate that speed with ICD-9-CM is not necessarily a predictor of speed with ICD-10-CM/PCS. For example, compare the following:

- Participant A's average ICD-9-CM coding time was 21.7 minutes, and this participant's average ICD-10-CM/PCS coding time was 42.7 minutes.
- Participants B and C's average ICD-9-CM coding time was between 18 and 19 minutes, and their average ICD-10-CM/PCS coding time was 32.7 minutes.

In this example, all three participants were only three minutes apart on their average ICD-9-CM coding time; however they were vastly different on their average ICD-10-CM/PCS coding time. Participant A's time variance was nearly double (almost a 100 percent increase), whereas participants B and C experienced approximately a 72 percent increase in the coding time. These data strongly suggest that no "perfect" ICD-10-CM/PCS coder exists because education and/or experience did not appear to have an impact.

However, the findings may reveal some hope for efforts to mitigate the expected loss of productivity. The negative correlation between productivity and quality reveals that, as the average time to code increases, the quality of the coding decreases. Coding managers should consider establishing a time limit for coding ICD-10-CM/PCS records, especially during the initial stages of the implementation. These results indicate that longer coding times do not result in higher quality.

The IRR results are not unexpected. We found that the IRR was higher for ICD-9-CM cases (group A) than for ICD-10-CM/PCS cases (group B). However, given the more than 30-year history of coding with ICD-9-CM and the lack of experience with ICD-10-CM/PCS, this finding is not unexpected. A more detailed analysis of the IRR results reveals that four cases (cases B3, B9, B15, and B24) have total agreement ($Pa = 1$) in ICD-10-CM, with three of the four (cases B3, B9, and B24) assigned an unspecified code. It was surprising that in just two of the cases (cases B3 and B9), similar results were found in the matched selected cases (cases A3 and A9). These cases must have been clear cut because the same result was found in the procedural coding for these two cases, even in ICD-10-PCS. Slightly more than half (14 of 27) of the cases coded in ICD-9-CM had total observable agreement. When the similar cases coded in ICD-10-CM were examined, two were found to have total agreement, three had high agreement ($Pa = 0.71$), five had moderate agreement ($Pa = 0.33–0.47$), and four had low agreement ($Pa = 0.09–0.14$). The investigation was continued by tracing back to check the codes for the cases with moderate and low agreement. One example of this in-depth analysis is found in case B14. All of the coders were convinced that the case was an ulcer with hemorrhage, but they were not sure whether the case was acute or chronic or an unspecific ulcer with hemorrhage. Another example is case B22; most of the coders considered this case to be non-ST segment elevation myocardial infarction, except for two coders, one of whom selected "other fluid overload" and the other of whom settled on "unspecified atherosclerosis." From this examination of IRR, we see that reaching substantial levels of agreement among ICD-10-CM/PCS coders, consistent with the levels achieved for ICD-9-CM, will take time and additional guidance from the controlling authorities.

Table 3 demonstrates that IRR, just like the average coding times, does vary by education and years of experience. Unfortunately, this sample size was not sufficient to reveal a meaningful pattern. For example, where the coder with the higher level of education might be expected to have a higher IRR with the ICD-10 gold standard coder, this coder did not. In fact, the coder with the high school education had the highest ICD-10-CM/PCS IRR. The findings for years of coding experience also showed no discernible pattern. The ICD-9-CM IRR results were even less revealing. No gold standard ICD-9-CM coder was identified; thus categories with only one participant could not have an IRR calculated.

The negative correlation of IRR with time to code a record strongly suggests that coders should not be allowed to spend an open-ended amount of time trying to code a record in ICD-10-CM/PCS. Once a predetermined time limit (set by the organization depending on the complexity of the cases) has been reached, the coder should seek guidance and assistance. These results indicate that longer amounts of time do not result in greater agreement or accuracy.

Study Limitations

As with any research, this time study has limitations. One limitation is that the data are based on an attempt to simulate real practice but do not in fact reflect actual coding production in the normal course of operations. In addition, the size of this study was limited to a total of 54 inpatient cases, in two groups of 27 cases each, and comparisons in the preliminary findings are between six participants, who coded all 54 cases. Furthermore, sample selection reflected the most common MS-DRGs at UC, but only a limited number of cases were included for each MS-DRG.

The results of this time study are further limited to making inferences on initial productivity losses and are not predictive of longer-term coding productivity with ICD-10-CM/PCS. A longitudinal study, with repeat coding by study participants after months of ICD-10-CM/PCS coding, would be needed to infer how long the initial productivity loss might last and predict what the new productivity expectation might be.

Further Research

Because this study focused on initial productivity losses, further study is needed to determine coding productivity changes over time. A follow-up study to determine the longer-term (six-month or more) impact of ICD-10-CM/PCS on coding productivity would be extremely informative during this important transition.

Conclusion

This study was designed to effectively simulate coding practice with ICD-9-CM and ICD-10-CM/PCS and reliably measure the time required to fully code an inpatient health record. The study supports the hypothesis that initial coding productivity using the ICD-10-CM/PCS classification systems will be significantly lower than current coding productivity using the ICD-9-CM classification system.

These findings indicate that previous estimates of initial coder productivity loss with ICD-10-CM/PCS may have been understated. Also, the study provides a strong indication of a significant return on investment for staff training time because study participants with the greatest amount of ICD-10-CM/PCS training experienced the lowest group average initial productivity loss at 54.4 percent. Finally, this study revealed that extensive time taken to code a record does not necessarily result in an increase in coding quality. These results are informative for ongoing planning to successfully navigate the transition to ICD-10-CM/PCS, particularly in budgeting for staff resources and determining ICD-10-CM/PCS training plans.

Mary H. Stanfill, MBI, RHIA, CCS, CCS-P, FAHIMA, is vice president of HIM consulting services at United Audit Systems, Inc., in Cincinnati, OH.

Kang Lin Hsieh, MS, is a PhD student in the School of Biomedical Informatics at the University of Texas Health Science Center at Houston in Houston, TX.

Kathleen Beal, MPA, CPHQ, RHIA, is division director of care management at the University Hospital of Cincinnati in Cincinnati, OH.

Susan H. Fenton, PhD, RHIA, FAHIMA, is an assistant professor and assistant dean for academic affairs in the School of Biomedical Informatics at the University of Texas Health Science Center at Houston in Houston, TX.

Notes

1. Hallowell, Bruce. "What Canada Can Teach the U.S. about ICD-10 Conversion." *Healthcare Informatics*, September 13, 2011. Available at http://www.healthcare-informatics.com/article/what-canada-can-teach-us-about-icd-10-conversion.
2. Innes, Kerry, Karen Peasley, and Rosemary Roberts. "Ten Down Under: Implementing ICD-10 in Australia." *Journal of AHIMA* 71, no. 1 (2000): 52–56.
3. IBM Corp. *IBM SPSS Statistics for Windows*. Version 20.0. Armonk, NY: IBM, 2011.
4. Microsoft Corp. *Microsoft Excel*. Version 2010. Redmond, WA: Microsoft, 2010.

Table 1

Study Approach to Elements Affecting Productivity

Elements Affecting Productivity	Study Approach
Complexity of the case (severity of illness)	Controlled: Two groups of inpatient cases were carefully constructed to be of similar complexity.
Health record format (paper, hybrid)	Controlled: The same health record format was used for all inpatient cases.
Coder familiarity with record format	Controlled: All coding professionals were initially unfamiliar with the University of Cincinnati health record format, and all received the same amount of orientation and training to access University of Cincinnati inpatient health records electronically.
Coding methodology (book, logic-based encoder, book-based encoder, computer-assisted coding)	Controlled: All study participants used the same encoder to code cases in both ICD-9-CM and ICD-10-CM/PCS. All study participants were previously familiar with the encoder used.
System access (downtime)	Controlled: While downtime is often a factor in real-time production coding, for the purposes of the study downtime was to be subtracted from total coding time. In actuality, none of the participants in the preliminary analysis reported any downtime or delays due to system access.
Sufficiency of health record documentation	Measured: Participants recorded narrative notes on each case describing any issues related to lack of clinical documentation.
Individual gender, age, knowledge, education, and reading speed	Measured: Participant descriptive information was captured, including gender, age, highest education level, professional credentials/certificates, years of experience coding with ICD-9-CM, amount of ICD-10-CM/PCS training, and extent of experience coding with ICD-10-CM/PCS.

Table 2

MS-DRGs Included in the Study

• 65	• 433, 345, 356
• 190, 191, 192	• 378
• 193, 194, 195	• 392
• 231-236	• 470
• 247	• 515-538
• 287	• 683
• 291, 292, 293	• 690
• 310	• 871
• 313	• 885

Table 3

Participant Differences

Participant Characteristic	Additional Time to Code with ICD-10-CM/PCS versus ICD-9-CM	ICD-9-CM Interrater Reliability (Non–Gold Standard)	ICD-10-CM/PCS Interrater Reliability
Highest education level achieved			
High school (1 participant)	18.66 minutes	----[a]	0.574
Associate degree (2 participants)	38.36 minutes	0.652	0.486
Bachelor's degree (2 participants)	27.49 minutes	0.728	0.436
Master's degree (1 participant)	17.82 minutes	----[a]	0.345
ICD-9-CM coding experience			
10 years or less (1 participant)	13.84 minutes	----[a]	0.496
11–15 years (1 participant)	20.06 minutes	----[a]	0.573
16–20 years (3 participants)	44.59 minutes	0.652	0.405
21+ years (1 participant)	27.78 minutes	----[a]	0.497

Note: Differences were averaged when a category had more than one coder.
[a] With only one participant and no ICD-9-CM gold standard coder, the interrater reliability could not be calculated for these categories.