

The Impact of Physician Quality Measures on the Coding Process

by Mark Morsch, MS; Ronald Sheffer, Jr., MA; Susan Glass, RHIT, CCS-P; Carol Stoyla, CLA; and Sean Perry

Abstract

Physician coding and billing is undergoing a major change that has expanded the responsibility of coders. The change is taking the form of Centers for Medicare and Medicaid Services (CMS) quality measures that are part of the new Physician Quality Reporting Initiative (PQRI), the successor to the Physician Voluntary Reporting Program (PVRP). The PQRI is a voluntary reporting program that provides a financial incentive for physician participation in the form of a 1.5 percent bonus payment for covered Medicare physician fee schedule services. This paper presents an analysis of the impact of the PQRI on three aspects of the coding process: (1) the frequency of occurrence of PQRI reportable cases, (2) the performance of a computer-assisted coding (CAC) system to assist the quality measure coding process, and (3) measurement of the coding effort required per case.

Why is the PQRI a major change for physician coding and billing? First, the mechanism for reporting the measures is the current claims processing system. The quality measures are reported using CPT (Current Procedural Terminology) category II codes either on CMS-1500 forms or electronic 837-P forms. Secondly, these measures can be complex, combining criteria from the patient demographics (gender and age), medical and surgical history, current and newly prescribed medications, diagnostic and laboratory tests, course of treatment, final diagnoses, content of documentation, and the timeline of events. Lastly, a single case may meet the criteria of zero, one, or more than one quality measures.

Introduction

In this paper we present an analysis of the impact of Physician Quality Reporting Initiative (PQRI) measures on the coding process for selected medical specialties. Using a large database of current physician notes, this impact is quantified by tabulating the frequency of occurrence for several quality measures in emergency medicine and radiology. Each quality measure has been analyzed for its typical case as well as the variations identifiable through special modifiers: 1P, 2P, 3P, and 8P. These modifiers indicate the reasons why the preferred protocol was not followed.

We also present the test results of a computer-assisted coding (CAC) system based on natural language processing (NLP) that has been developed to assign the quality measure codes. The level of agreement is reported when comparing the CAC system output to the changes made by human coders. The percentage agreement is a gauge of the accuracy of the CAC system, although several factors come into play when analyzing the results, such as the coder learning curve for new guidelines, the voluntary nature of the program, and the data extrinsic to the CAC system that influence quality measure assignment.

A third component of the impact of the PQRI on the coding process is the time required to document, collect, and codify the measures. In this study we report the timing data tracked during the coding process for cases with quality measures. These results are compared to similar cases coded prior to the implementation of the PQRI. This analysis quantifies the impact of the collection and coding of quality measures on the coder workload.

Background

The final revisions to the PQRI guidelines were published by the Centers for Medicare and Medicaid Services in mid-June 2007 for implementation beginning with date of service July 1, 2007.¹ The PQRI guidelines define 74 measures across multiple medical specialties. An individual physician working in a single specialty will typically report on a handful of measures that specifically apply to his or her medical specialty area. This paper includes an analysis of the emergency medicine and radiology specialty areas.

The LifeCode engine² is the CAC technology used in this analysis to assist the quality measure coding process. In commercial use since 1998, LifeCode is an NLP-based computerized coding engine. To support the extraction of the PQRI measures, the LifeCode engine was enhanced to recognize and codify the data elements specified in the PQRI guidelines. Most of the measures are codified using CPT category II codes, which are five characters long, consisting of four digits followed by one alphabetic character.

A total of nine measures are covered in this study: two measures applicable to radiology and seven measures applicable to emergency medicine. The seven emergency medicine measures were those developed with input from the American College of Emergency Physicians (ACEP). Table 1 lists the nine measures, their associated medical specialties, and the related CPT category II codes.

Analysis

The data for this study are from selected users of the LifeCode engine during a three-week period in July 2007. Users were selected based on the criteria of coding any of the nine PQRI measures for PQRI-eligible claims during the designated time period. Users that did not code any of the PQRI codes during the month of July were not included in this study.

The final data covered 2,632 radiology cases from 179 radiology facilities and 569 emergency medicine cases from 13 emergency medicine facilities. The statistics presented in the results section describe these data, along with productivity data from earlier months for purposes of comparison.

Results

The first subject of the analysis is the frequency of PQRI cases for emergency medicine and radiology. PQRI case percentage varies based upon the number of measures that apply to the medical specialty, the case mix of the facility, and the measures selected for reporting by the provider or facility. Figure 1 shows the PQRI case percentage for the 13 emergency medicine facilities. The average PQRI case percentage was 1.86 percent, with a low of 0.28 percent and a high of 6.78 percent.

Figure 2 shows the PQRI case percentage for the 179 radiology facilities. The average PQRI case percentage was 0.40 percent, with a low of 0.01 percent and a high of 7.20 percent. These case percentage averages are consistent with the relative number of measures defined for emergency medicine versus radiology.

The second aspect of the analysis is the accuracy of CAC technology in assigning the quality measure codes. Coders reviewed the output of the CAC system and made edits to the PQRI codes consistent with their individual judgment. The work was performed as part of routine production coding operations for each of the facilities. The CAC system stores the coding edits for each case, allowing a comparison between CAC output and the final codes assigned by the users.

System accuracy is calculated using an agreement rate that is the percentage of final PQRI codes assigned by the coders that matched the codes produced by the CAC system. This is the same formula as a recall statistic.³ Agreement rate was calculated two ways: (1) as the percentage of matching CPT category II codes and (2) as the percentage of matching CPT category II codes plus modifiers. Figure 3 shows the agreement rates for the emergency medicine facilities. Each facility is shown with a pair of statistics: the upper dark bar of each pair represents the agreement rate of CPT II codes plus modifiers and the lower light bar represents the agreement rate of CPT II codes. Overall, the agreement rate was 80.6 percent for CPT II codes plus modifiers and 89.0 percent for CPT II codes alone.

The agreement rates for the radiology cases are shown in Figure 4. In this chart, the rates have been grouped into five levels, and the bars indicate the number of facilities that fall into each level of agreement rate. The dark-colored bars indicate the agreement rate for CPT II codes plus modifiers, and the light-colored bars represent the agreement rate of CPT II codes. For all radiology facilities in the study, the agreement rate was 61.9 percent for CPT II codes plus modifiers and 71.3 percent for CPT II codes alone.

The third aspect of the study is measuring the effect of PQRI coding on coder productivity. The CAC system stores the amount of time a coder spends reviewing each case. Figure 5 shows the average time spent per case for the July cases as compared to the same types of cases over the previous three months. The lightest-colored bars are the PQRI cases. The time impact for the radiology cases is dramatic; the average time per case increased by 277 percent from 58 seconds to 220 seconds. The impact on emergency medicine cases was much less, with an increase of 13 percent over the previous three-month average, from 157 seconds to 177 seconds.

Discussion

Anecdotally, the feedback from users has been very positive concerning the performance of the CAC system for PQRI coding. It's important to consider several factors when interpreting these statistics. First, the PQRI program is voluntary, and providers have some options in selecting which measures to collect and report. This study measured changes but did not collect the reasons behind the changes. Changes made based on CAC system errors were not distinguishable from changes made based on other information, such as the choice not to report certain cases or the availability of supplemental documentation to the coder that was not available to the CAC system. This made it difficult to compute a level of precision for the CAC system and to quantify false positive results. Second, because it was early in the reporting period, some facilities were still deciding whether and to what extent to participate in the program. Coding policies and practices at individual facilities were evolving during the period. Lastly, many coders were still learning the guidelines, so it is natural to expect a lower level of human coder accuracy and consistency this early in the process. That said, we believe this study provides valuable information in regard to the effects of implementing the PQRI guidelines and how a CAC system performed during the initial implementation.

Mark Morsch, MS, is the vice president of NLP and software engineering at A-Life Medical, Inc., in San Diego, CA.

Ronald Sheffer, Jr., MA, is the manager of NLP development at A-Life Medical, Inc., in San Diego, CA.

Susan R. Glass, RHIT, CCS-P, is a senior QA specialist at A-Life Medical, Inc., in San Diego, CA.

Carol Stoyla, CLA, is the director of compliance, coding, and software QA at A-Life Medical, Inc., in San Diego, CA.

Sean Perry is a research linguist at A-Life Medical, Inc., in San Diego, CA.

Notes

1. Centers for Medicare and Medicaid Services. 2007 Physician Quality Reporting Initiative Specifications Document. Available at http://www.cms.hhs.gov/PQRI/downloads/Measure_Specifications_061807.pdf
2. Heinze, Daniel, et al. "LifeCode: A Deployed Application for Automated Medical Coding." *AI Magazine* 22, no. 2 (2001): 76–88. Available at <http://www.alifemedical.com/documents/LifeCodeAIMagazine.pdf>.
3. Wikipedia Foundation. Definitions of Recall and Precision. Available at http://en.wikipedia.org/wiki/Recall_%28information_retrieval%29.

Table 1

Selected PQRI Measures

Measure Number	Measure Name	CPT II Codes
10 – Rad	Stroke and Stroke Rehabilitation: Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) Reports	3110F, 3111F, 3112F
11 – Rad	Stroke and Stroke Rehabilitation: Carotid Imaging Reports	3100F
28 – EM	Aspirin at Arrival for Acute Myocardial Infarction (AMI)	4084F
54 – EM	Electrocardiogram Performed for Non-Traumatic Chest Pain	3120F
55 – EM	Electrocardiogram Performed for Syncope	3120F
56 – EM	Vital Signs for Community-Acquired Bacterial Pneumonia	2010F
57 – EM	Assessment of Oxygen Saturation for Community-Acquired Bacterial Pneumonia	3028F
58 – EM	Assessment of Mental Status for Community-Acquired Bacterial Pneumonia	2014F
59 – EM	Empiric Antibiotic for Community-Acquired Bacterial Pneumonia	4045F

Figure 1

Emergency Medicine PQRI Case Frequency

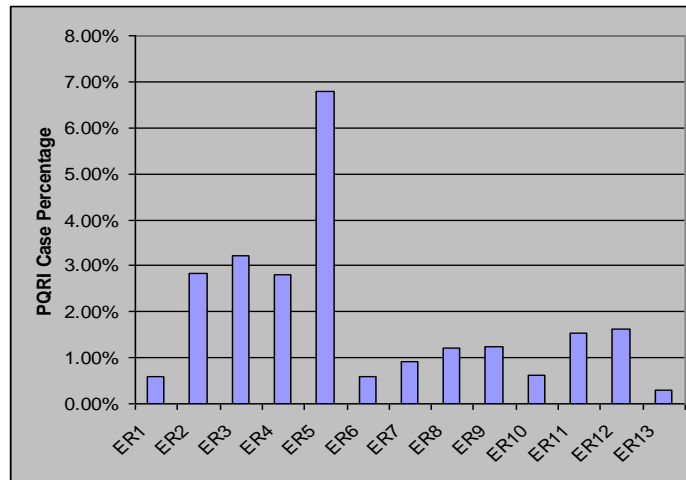


Figure 2

Radiology PQRI Case Frequency

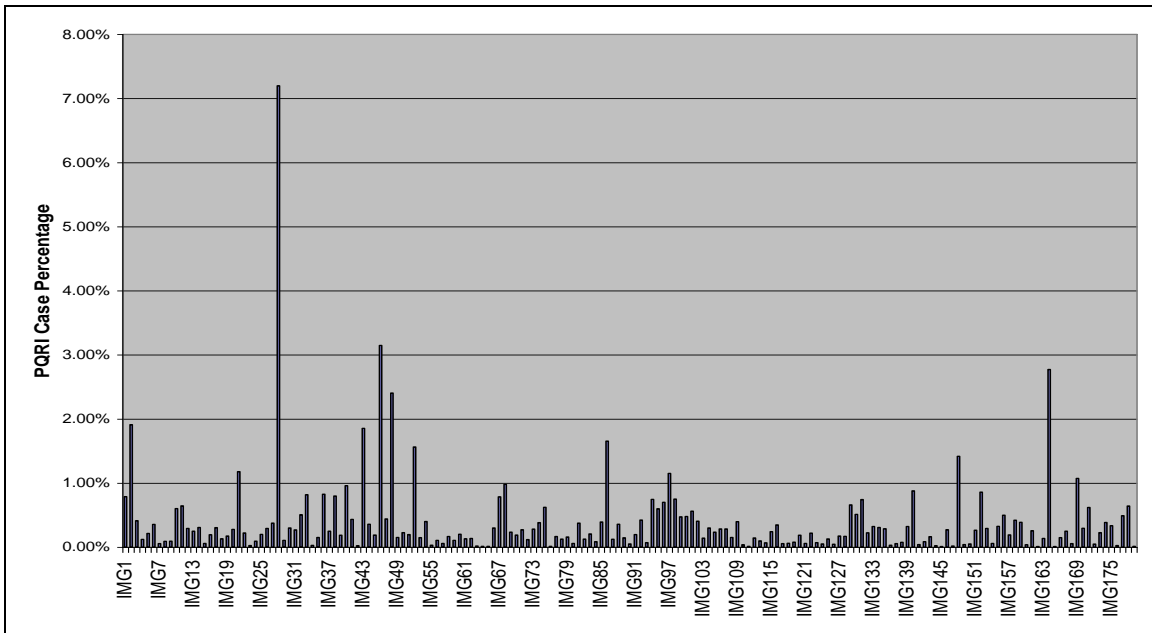


Figure 3

Emergency Medicine Agreement Rates

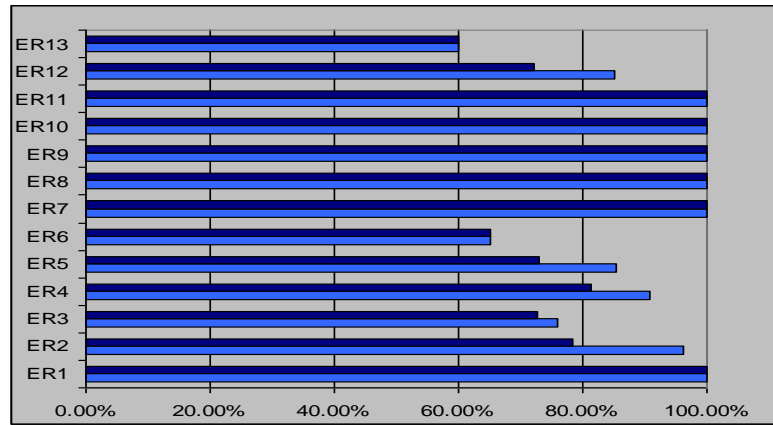


Figure 4

Radiology Agreement Rates

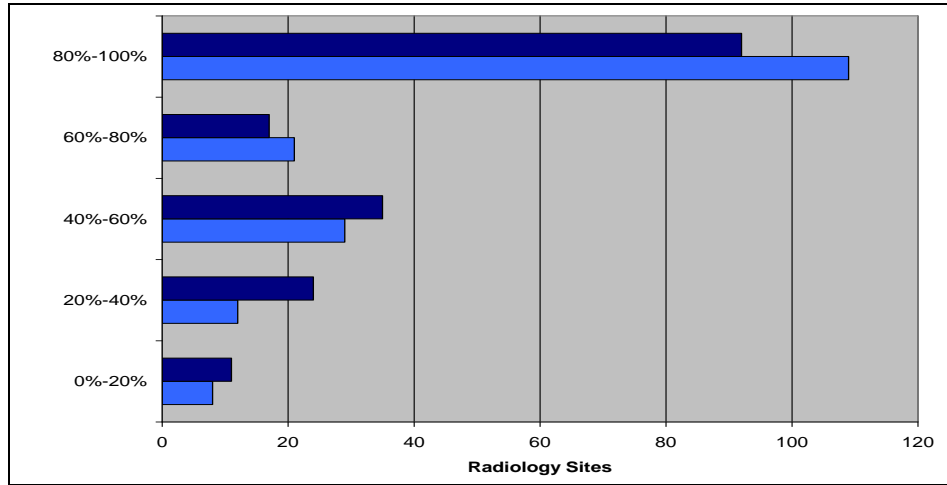


Figure 5

Coder Productivity

