Coding Verbatim Responses Using an Auto-coding Program Based on a Two-step Matching Process: National Hospital Ambulatory Medical Care Survey Emergency Department Data, 2015

Akintunde Akinseye, Brian W. Ward
Division of Health Care Statistics, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference

Disclaimer
The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics or the Centers for Disease Control and Prevention.

Introduction
The importance of using standardized medical coding systems to study medical diagnoses, symptoms, and diseases was recognized over fifty years ago (Bain & Spaulding 1967), and is perhaps even more important today (Borman 2017). Currently a number of medical coding classification systems exist, such as the International Classification of Diseases, 10th revision, Clinical Modification (ICD-10-CM), the International Classification of Diseases, 10th revision, Procedural Coding System (ICD-10-PCS), Current Procedural Terminology (CPT), and Healthcare Common Procedure Coding System (HCPCS). These classification systems provide researchers with a mechanism to quantifiably examine phenomena such as diagnoses and performance of medical procedures in meaningful ways that inform the epidemiology, prevention, and management of diseases. However, these systems are not uniformly applied across different electronic health record systems, and even when they are, there may be value in having access to the original written notes (Rosenbloom et al. 2011). To use these original written notes for analytical purposes, researchers are faced with the task of assigning a standardized medical coding classification system to the free text obtained from large health care data sources.

The National Hospital Ambulatory Medical Care Survey (NHAMCS) is conducted by the National Center for Health Statistics (NCHS) and has been fielded annually since 1992. NHAMCS is a nationally-representative survey of nonfederal general and short-stay hospitals that can be used to assess the utilization of ambulatory medical care services in hospital emergency departments (ED) and other settings. Data are manually abstracted from medical records using a computerized questionnaire. During the NHAMCS data collection process, a variety of data are collected in free-text format (verbatim), including medical diagnoses, procedures/labs, type of injury, medications ordered/prescribed, and patient’s reason for visit. After data collection, certified medical coding professionals are hired to manually code these verbatim data elements according to specific medical coding classification systems. This coding process is costly, and decreases the timeliness of NHAMCS data release to the public. The increasing cost of the manual coding of these data and the continued push to improve the timeliness of NHAMCS has resulted in a need to seek alternative methods for coding the verbatim data elements collected in the survey.

The main objective of this study was to develop a model to explore a more efficient approach to coding verbatim data elements from NHAMCS using the staff and resources currently available at NCHS. In the development of this approach, we focus specifically on a patient’s reason for visit (RFV) to the ED, and use already-coded ED data from the 2015 NHAMCS to assess the accuracy of this proposed approach.

Code Assignment Process
The process we used for coding NHAMCS data comprised three main stages: percent matching, direct matching, and manual review (Figure 1). The first two stages were implemented using SAS software, while the third stage, which is still underway, requires manual coding by certified medical coders. This entire process is dependent on the RFV data dictionary – a coding classification system that was created by NCHS for RFV notes. These notes have been broadly categorized into 9 modules (see NCHS 2018a, Appendix II), with each module pertaining to a certain group of patient complaints. For example, Module 1 pertains to the symptoms reported by the patient, and Module 3 pertains to treatments mentioned in the patient’s notes. The distinction between each module was helpful in categorizing the data set, so each record in the 2015 NHAMCS ED (uncoded raw data) was grouped into one of the 9 categories. This stage was done as a time saving measure while processing the data so instead of running all of the data against each of the categories of the RFV data dictionary, only the portions of the dictionary predetermined to be associated with any given verbatim text in the data would be cross-examined.

Figure 1. Proposed three-stage model for coding verbatim data entries in the National Hospital Ambulatory Medical Care Survey Emergency Department data
Prior to implementing the three main stages mentioned above, the un-coded verbatim data were optimized to remove typos, expand abbreviations, and delete excess words. The main purpose of data optimization was to improve the likelihood of the raw verbatim data being successfully assigned a RFV code for the percent matching stage. Typos were fixed using SAS function `COMPGED` by creating a words matrix that pulled together similar-sounding words. The resulting list of matches were between typos and similar-sounding but correctly-spelled words from the RFV data dictionary, which were then reviewed manually and the typos corrected by the algorithm. In trying to expand the abbreviations, we considered the significant variability in the use of medical abbreviations, and so limited our list to commonly used abbreviations. Finally, excess words were deleted based on our definition that any word not found in the RFV data dictionary was ‘junk’. By compiling all the words in the RFV data dictionary into a macro variable, that variable was used to strip the raw uncoded data to only words that had been found in the data dictionary. The deleted words were also excluded from the percent matching calculations.

After the data optimization steps, the percent matching scores were calculated. During this stage, each verbatim entry was compared, word-for-word, with all of the code descriptions within the module to which the verbatim data had been previously categorized, and matching scores were calculated. Of the matching scores calculated, the code with the highest value was attributed to the verbatim entry. This score was calculated by dividing the number of words matched between the verbatim data and the RFV dictionary by the total number of words found in a verbatim and RFV dictionary combined:

\[
\text{\% Word Matching Score} = \frac{\text{\# Words Matched between Verbatim Entry and RFV Dictionary}}{\text{Total \# of Words Found for a Verbatim Entry and RFV Dictionary}} \times 100
\]

For example, if a verbatim RFV entry was “headache” (1 word) was compared with the following three RFV data dictionary entries: “migraine”, “headache”, or “painful headache,” the resulting percent word matching scores would be 0.0%, 100.0%, and 66.7%, respectively. Being that the second RFV data dictionary entry had the highest matching score, its corresponding RFV code was assigned to the verbatim entry. Thereafter a threshold for the percent matching was set, such that all percentages that did not meet this criterion were not assigned a code. For this study thresholds were set to greater than 0% (any word matching), 50%, 80%, and 90%. Also, where the percent word matching score for two or more RFV codes was equal, multiple codes were assigned to the verbatim entry. As part of the manual review stage, further scrutiny by a certified medical coder would be needed to select the most appropriate of the multiple codes that were assigned.

For the following stage, the 2015 NHAMCS ED verbatim data with previously assigned percentage matched codes, was run through the direct matching stage portion of the algorithm. The algorithm was designed to ensure that if a verbatim entry was assigned a code during the percent matching stage, that code was not overwritten during the direct matching stage. This stage involved using a list of verbatim entries that had already been coded in previous years of NHAMCS (in this study, 2013-2014...
ED data) to assign RFV codes to the 2015 NHAMCS ED verbatim data, but only if the entries are exact matches. One common occurrence in coding the RFV notes is that certain entries result in more than one code (multi-code verbatim entries). This analysis was limited to single-code verbatim entries to avoid attributing unrelated codes to a 2015 NHAMCS ED verbatim entry, and in so doing introduce error.

For the third stage, which is still underway, manual review would be conducted to compare the medical-coder assigned 2015 NHAMCS ED data to the algorithm assigned 2015 NHAMCS ED data, to identify areas of disagreement as a result of miscoded or uncoded verbatim data. In future iterations of this algorithm, being that there won’t be any reference data sets to make comparisons to, manual coding would be performed by trained medical coders on any remaining uncoded/miscoded verbatim entries. These manually-reviewed verbatim entries would then be reintroduced to the RFV dictionary, thereby making the entire algorithm more robust for coding future years of data using the same model.

Results from Empirical Assessment of the Model

Our empirical examination of this model tested the coverage and accuracy of the percent and direct matching stages combined, but not the manual review stage. To achieve this, we compared the codes assigned by the model to already-coded NHAMCS ED visit data of the same year. The 2015 NHAMCS ED data was collected for 243 hospitals, for a total of 21,061 visit records (approximately 137 million ED visits [weighted]) (Hall et al. 2018). Additional information on the methodology of the 2015 NHAMCS is available elsewhere (NCHS 2018a, 2018b). For each visit in the NHAMCS ED dataset, there is a potential for up to five separate RFVs to be entered. Of the 21,061 visits in the 2015 NHAMCS ED data, there were a total of 43,565 RFV verbatim entries collected, all of which had previously been coded by certified medical coders using NCHS’ RFV coding classification system (see NCHS 2018a, Appendix II).

Below are the results from the initial assessment, shown using different thresholds, with both the percent matching and direct matching stages (Table 1). The direct matching results, which are included in the table below, were all 100% being that they were exact matches. When running the model for any percent word matching threshold (that is >0%), 39,318 of the 43,565 RFV verbatim entries (or 90.3%) resulted in the assignment of a medical code. While this coverage was high, the accuracy of those RFV verbatim entries coded was 78.2%.

Table 1. Coverage and accuracy of proposed model applied to reason for visit verbatim entries (n=43,565): National Hospital Ambulatory Medical Care Survey, 2015 Emergency Department data

<table>
<thead>
<tr>
<th>% Matching Threshold</th>
<th>Coverage</th>
<th>Accurate</th>
<th>Inaccurate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
</tr>
<tr>
<td>&gt;0% (Any)</td>
<td>39,318</td>
<td>90.3</td>
<td>30,750</td>
</tr>
<tr>
<td>≥50%</td>
<td>36,456</td>
<td>83.7</td>
<td>29,813</td>
</tr>
<tr>
<td>≥80%</td>
<td>27,970</td>
<td>64.2</td>
<td>25,005</td>
</tr>
<tr>
<td>≥90%</td>
<td>24,487</td>
<td>56.2</td>
<td>22,421</td>
</tr>
</tbody>
</table>

As the word matching threshold increased, a decrease in coverage and an increase in accuracy was observed. The biggest difference was found between moving the word matching threshold from ≥50% to ≥80%, where coverage decreased by 19.5 percentage points (from 83.7% to 64.2%) and accuracy increased by 7.6 percentage points (from 81.8% to 89.4%). At the ≥90% threshold, a little over half of the RFV verbatim entries were auto-coded (56.2%), with an accuracy of 91.6%. The accuracy and coverage remained unchanged at higher matching thresholds.

Next Steps

In future iterations of the algorithm, we will be exploring ways to improve the coverage, but more importantly the accuracy. We plan to review the uncoded and miscoded verbatim entries to determine if there are any patterns not being picked up by the algorithm. We also plan on enriching the RFV data dictionary with any verbatims that were reviewed in the 3rd stage of the model. We plan to handle the excess words differently. Instead of deleting words, and in so doing lose context, we will explore the use of weights whereby words found in the RFV data dictionary would be more heavily weighted than words not found in the dictionary.

In this iteration of the algorithm, we restricted assignment of codes at the direct matching stage so that already assigned percent matching codes were not overwritten. We plan on exploring changes to the accuracy of the model when the direct matching stage is unrestricted.

To improve the results of the direct matching stage, we plan to include more years, say 10 years, of already-coded data as opposed to just 2013 and 2014 data that was used for the direct matching stage for this study. Also, previously coded RFV data from the related National Medical Ambulatory Care Survey (NAMCS) could be employed to improve the RFV data dictionary.
Conclusion
In this study we described a three-stage model for coding data elements from NHAMCS Reasons for Visit data, using only computer software and the resources currently available at NCHS. With reliance on auto-coding in two of the model’s stages, there is indication for a more efficient code assignment of some of the verbatim data elements found in NHAMCS ED data. A preliminary empirical assessment of the combined percent and direct matching stages of this model provided support for its potential, especially as more stringent restrictions were placed on the word matching threshold. These restrictions resulted in a 91.6% accuracy of medical codes assigned to RFV. Part of the reason we will be exploring avenues to improve the accuracy of the model is to attempt to meet the 95% accuracy threshold to which medical coding contractors are currently held.

While accuracy is arguably most important, coverage should also be considered to assess the utility of the proposed model. With the most stringent restrictions placed on the model, a little over half (56.2%) of RFV verbatim entries were assigned a code. The eventual model is not intended to replace the contributions of certified medical coders, but our hope is that it will decrease the overall cost of medical coding and improve the timeliness of NHAMCS data release.

Implementation of this automated coding process has potential to result in a more efficient coding of verbatim data elements collected in NHAMCS, such as patient RFV, using resources currently available to NCHS. The results from this preliminary assessment yields an encouraging accuracy and coverage that merit more exploration. This model has the potential to be applied to other NHAMCS data elements collected in verbatim format, and even to other health care surveys conducted by NCHS, such as NAMCS. Its use could ultimately save financial and human resources, and has the potential for more timely release of NHAMCS data.

References


