

PROGRAM: ANALYTICS

GRADUATE PROJECT

ICD-10 CODING APPLICATION AND ANALYSIS

LAUREN TABER

GRAD 699

4/24/2014

Table of Contents

[ABSTRACT: 3](#_Toc417419987)

[KEY DEFINITIONS: 3](#_Toc417419988)

[INTRODUCTION: 3](#_Toc417419989)

[LITERATURE REVIEW: 4](#_Toc417419990)

[DATA EXPLORATION 5](#_Toc417419991)

[DATA DESCRIPTION 5](#_Toc417419992)

[RELATIONSHIP MAPPING 6](#_Toc417419993)

[ADDITIONAL VARIABLES 6](#_Toc417419994)

[APPLICATION DEVELOPMENT 7](#_Toc417419995)

[IMPEDIMENTS 10](#_Toc417419996)

[EXTENSION 10](#_Toc417419997)

[CITED LITERATURE: 11](#_Toc417419998)

# ABSTRACT:

Medical coding has been a relatively stable profession for many years with only minor updates as new advances occur in the medical field. In October 2015, an entirely new coding structure is being introduced, which will more than triple the number of codes available to describe medical procedures and diagnoses. Due to this significant change and the challenges it presents to the medical coding field, the purpose of this project is to design a system that explores possibilities of storing and interacting with the data in an unstructured database. This will be accomplished by setting up a graph database and creating a user interface to run queries against the data.

# KEY DEFINITIONS:

**Medical Coding:**

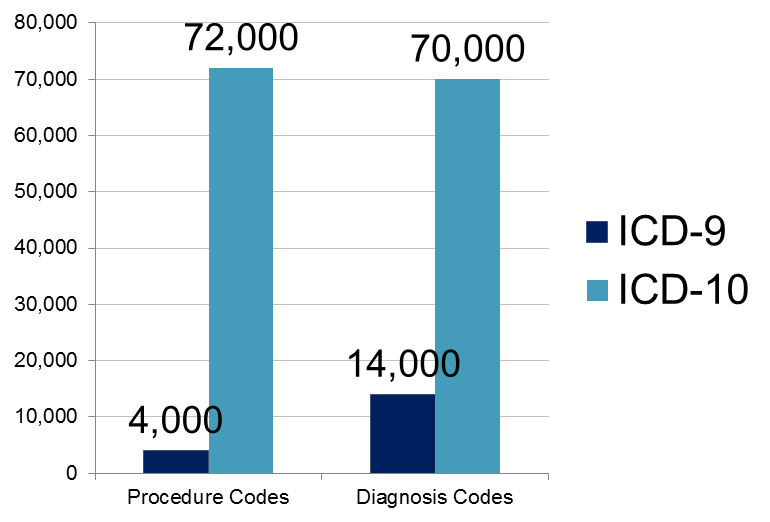
“Every time a patient receives professional health care in a physician’s office, hospital outpatient facility or ambulatory surgical center (ASC), the provider must document the services provided. The medical coder will abstract the information from the documentation, assign the appropriate codes, and create a claim to be paid, whether by a commercial payer, the patient, or CMS” (AAPC 2014).

**International Classification of Diseases (ICD):**

ICD is used “to classify diseases and other health problems recorded on many types of health and vital records, including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States. Finally, ICD is used for reimbursement and resource allocation decision-making by countries” (World Health Organization, 2014).

# INTRODUCTION:

Medical coding is the key mechanism used for processing health insurance claims. While there is a comprehensive list of codes describing all varieties of medical procedures and diagnoses, it is the job of medical coders to translate information from a patients chart into codes used for billing and claims processing. People interested in becoming medical coders go through fairly rigorous training designed to teach them to identify appropriate codes based on human anatomy, medical terminology, treatments, and procedures.

In the upcoming set of codes to be published in 2015, there are approximately 72,000 procedure codes and 70,000 diagnosis codes. The previous version had about 4,000 procedure codes and 14,000 diagnosis codes. The reason for this switch is that the medical field has changed significantly in the past 30 years since the ICD-9 codes were introduced. “The uses being made of coded data today go well beyond the purposes for which ICD-9-CM was designed” (Bowman, 2008). Clearly medical coding is becoming much more demanding and takes much more work to determine the correct code. To help minimize the stress this new influx of codes is putting on the medical coders, there needs to be a system that includes features that aid users in selecting the appropriate codes. It should also be designed to handle the increased load on the processing system caused by the significant increase in data.

# LITERATURE REVIEW:

Historically, medical codes have been used primarily for patient billing; however, “the system cannot adequately accommodate dramatic advances in medicine and medical terminology” (Bowman, 2008). The World Health Organization (WHO) is currently rolling out a new set of medical codes (ICD-10) that are much more specific and will result in “better documentation of disease states and complications, categorization of disease by community, and more specific tracking of healthcare outcomes” Rahmathulla (2014).

The CDC website also describes a number of benefits that will result from the switch to the ICD-10 codes.

ICD-10-CM/PCS code sets will enhance the quality of data for:

•Tracking public health conditions (complications, anatomical location)

•Improved data for epidemiological research (severity of illness, co-morbidities)

•Measuring outcomes and care provided to patients

•Making clinical decisions

•Identifying fraud and abuse

•Designing payment systems/processing claims

(CDC/National Center for Health Statistics, 2013)

Although the potential benefits are clearly significant and necessitate this major transition, there are a number of challenges that will be faced during the transition period. Specifically, medical coders face substantial challenges in transitioning to the new coding system. “Some hospital officials and physician leaders fear a decline in coder productivity, which could result in revenue losses from bad documentation, delays in inputting data and rejected claims because of improper coding or unproven IT systems” (Conn,2013).

The date that all hospitals will be required to make the switch over to the ICD-10 codes has been postponed several times. The most recent date scheduled is October 2015. These delays have had some pros and cons. On the positive side, it is giving people more time to train workers and update their IT systems. Conversely, some hospitals and clinics have spent quite a bit of money on retraining their workers only to have them continue using the same system for another year. All of this confusion only increases the probability of errors once the switch finally occurs.

According to Rahmathulla (2014), “Success for organizations during the transition is initially highly dependent on correct coding. Inaccurate coding would mean denial of reimbursements, prolonged periods of scrutiny into the healthcare records and downstream processes”. The key to moving forward is going to be finding a way to minimize the amount of coding errors by providing intelligent, user-friendly systems to aid medical coders as they adjust to the new ICD-10 codes.

# DATA EXPLORATION

The first part of this project is a detailed exploration of the dataset. Based on the initial analysis of the data, it can be appropriately imported into a database and manipulated to be most useful for the application to be developed. I have divided this process into 3 steps. The first step is to identify the data types and significance of each column. The second step is to determine what, if any, relationships exist between each data point. Finally, additional fields will be added to account for any supplementary variables that will be useful in creating relationships between data points.

## DATA DESCRIPTION

The ICD-10 code dataset contains two categories of data. In addition to the procedure codes and information, the dataset contains “headers” which represent the first three digits of the seven digit code. The two categories of data are distinguished by a binary flag with 0 representing a valid code and 1 representing a header. Included in the data file are four columns that I have labeled Code, Valid, Short Description, and Long Description.

*First 3 Database Entries Displayed As Table*

|  |  |  |  |
| --- | --- | --- | --- |
| Code | Valid | Short Description | Long Description |
| 001 | 0 | Central Nervous System, Bypass | Central Nervous System, Bypass |
| 0016070 | 1 | Bypass Cereb Vent to Nasophar with Autol Sub, Open Approach | Bypass Cerebral Ventricle to Nasopharynx with Autologous Tissue Substitute, Open Approach |
| 0016071 | 1 | Bypass Cereb Vent to Mastoid Sinus w Autol Sub, Open | Bypass Cerebral Ventricle to Mastoid Sinus with Autologous Tissue Substitute, Open Approach |

The table below contains a description of the conditions that are currently enforced on each data value in the dataset. Any future updates to the ICD-10 codes will be in the same format.

*Data Value Conditions*

|  |  |  |
| --- | --- | --- |
| Data Value | Data Type | Maximum Characters |
| Code | Integer | 7 |
| Valid | Binary | 1 |
| Short Description | String | 60 |
| Long Description | String | 400 |

## RELATIONSHIP MAPPING

There are two relationships that were easily identifiable from a sampling of the data set. The most obvious one is the relationship between the headers and the codes that begin with the code given to the header. This relationship is essentially a parent-child relationship (ie. 001 is a parent of 0016070). The second relationship is directly related to this parent-child relationship in that the description of the parent code explicates the anatomical location in which the procedure is being performed. Since the location is a substring of the long description, it is necessary to extract it so that it will be easier to work with later on.

A third relationship to note but was not used in the initial development of the application is related to the fifth digit of the code. This digit represents the approach taken during the procedure. There are five distinct approaches referenced within the data set. While this relationship exists, it was not determined to be significantly useful for the identification of desired codes by users.

## ADDITIONAL VARIABLES

As mentioned in the previous section, the location variable should be more clearly identifiable within the dataset in order to maximize the efficiency of queries against this key. Thus, a new column, Location, will be added to the file to represent this new variable.

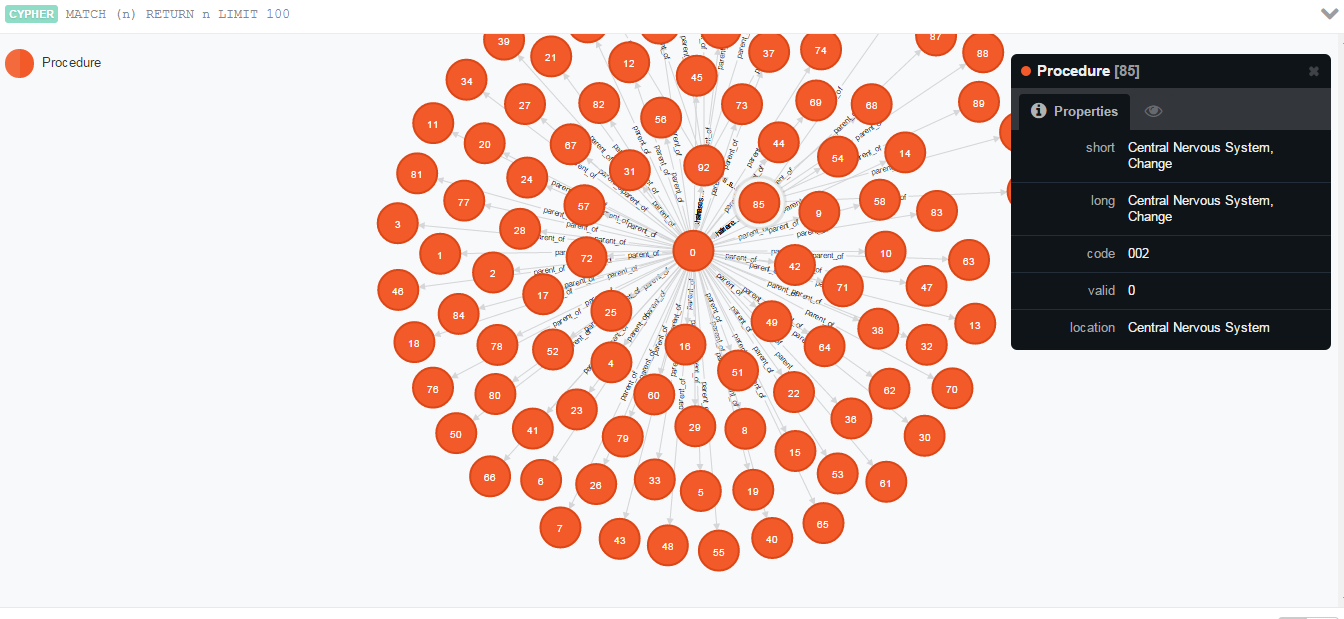
*First 3 Database Entries Displayed As Table (New Variable)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Code | Valid | Short Description | Long Description | Location |
| 001 | 0 | Central Nervous System, Bypass | Central Nervous System, Bypass | Central Nervous System |
| 0016070 | 1 | Bypass Cereb Vent to Nasophar with Autol Sub, Open Approach | Bypass Cerebral Ventricle to Nasopharynx with Autologous Tissue Substitute, Open Approach | Central Nervous System |
| 0016071 | 1 | Bypass Cereb Vent to Mastoid Sinus w Autol Sub, Open | Bypass Cerebral Ventricle to Mastoid Sinus with Autologous Tissue Substitute, Open Approach | Central Nervous System |

# APPLICATION DEVELOPMENT

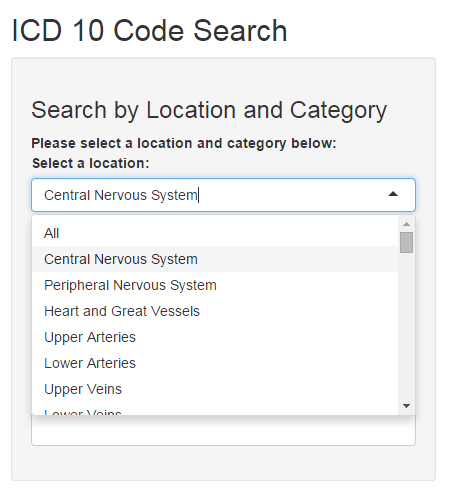
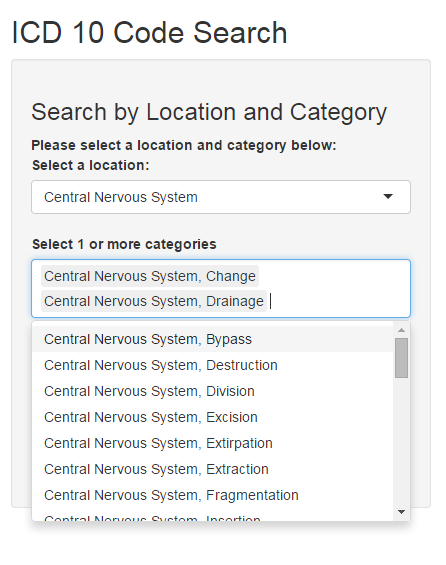
The source data file was a text file, which was imported into Microsoft Excel for the data exploration phase of the project. The application development phase began with importing the csv file from Excel into a Neo4j Graph Database. Each data point is a Procedure and has labels for Code, Valid, Short Description, Long Description, and Location. Each node is automatically provided with an index.

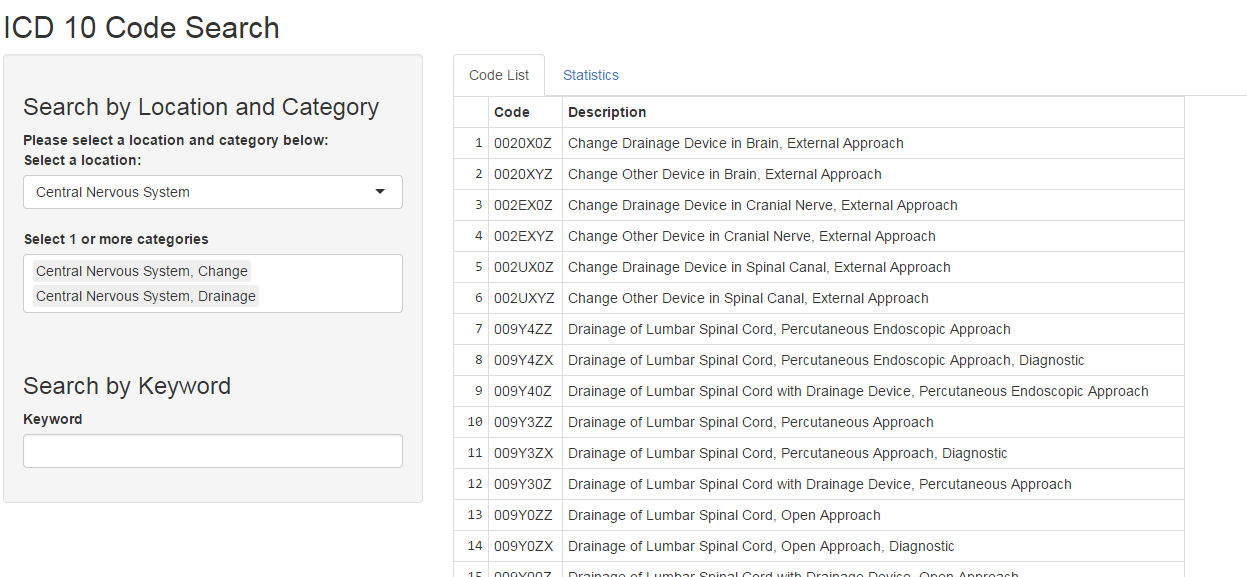
The next stage was to create the parent-child relationship identified above. This was done using Cypher directly in the Neo4j Database user interface. A small sampling of the resulting graph database, including relationships is show below. (Sample size: 100)



Now that the data has been imported and relationships have been added. The user interface can be developed. For this I have chosen to use RStudio and Shiny Analytics. The application consists of 2 search features and 2 results lists. The first search feature is a dynamic search that allows a user to select from a list of locations. The locations are the distinct list of location labels from the database. There is also an option to select “All” if the user would like to bypass the first filter criteria. Based on the user’s first choice, a second dropdown will appear with a list of categories. The user may choose one or more of these categories to populate the results tabs. The results will not be populated for this filter unless a category is chosen. The categories are populated by selecting header nodes that have a label consistent with the Location selection made by the user. If the user selects “All” from the Location dropdown, all header nodes will be displayed in the Category drop down. It should also be noted that the results set will only contain valid ICD-10 codes. Codes associated with headers will not be displayed.

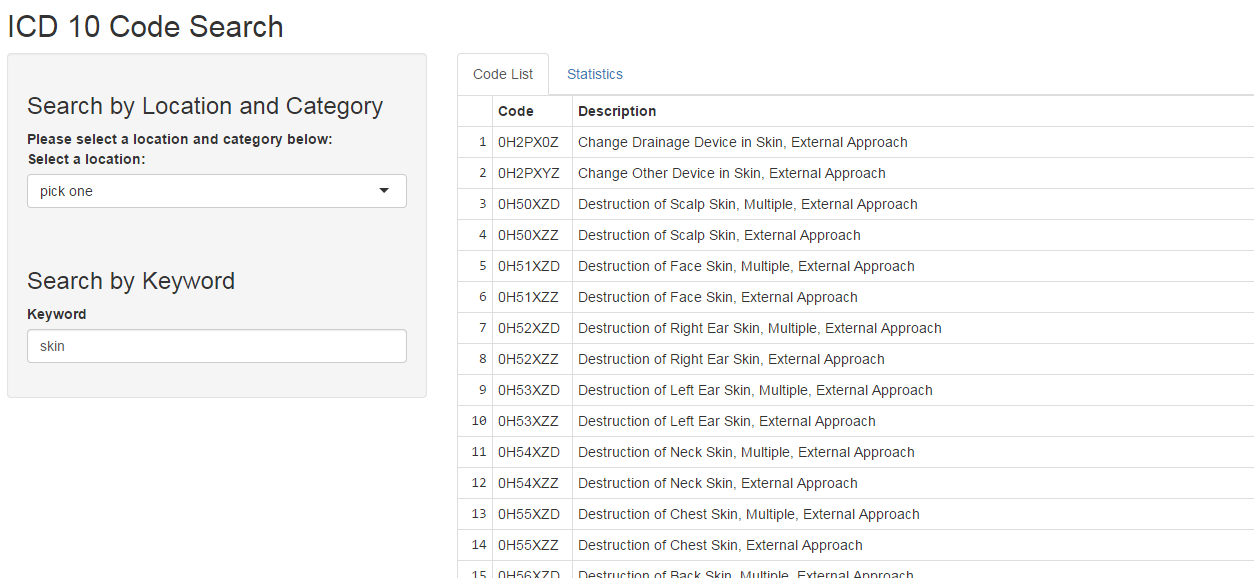
*Filter Selection*

*Result Set*

The second search feature is a simple keyword search. It is case insensitive and searches for the substring within the Long Description of each node. It does not take into consideration the location or category selections made in the other search feature.

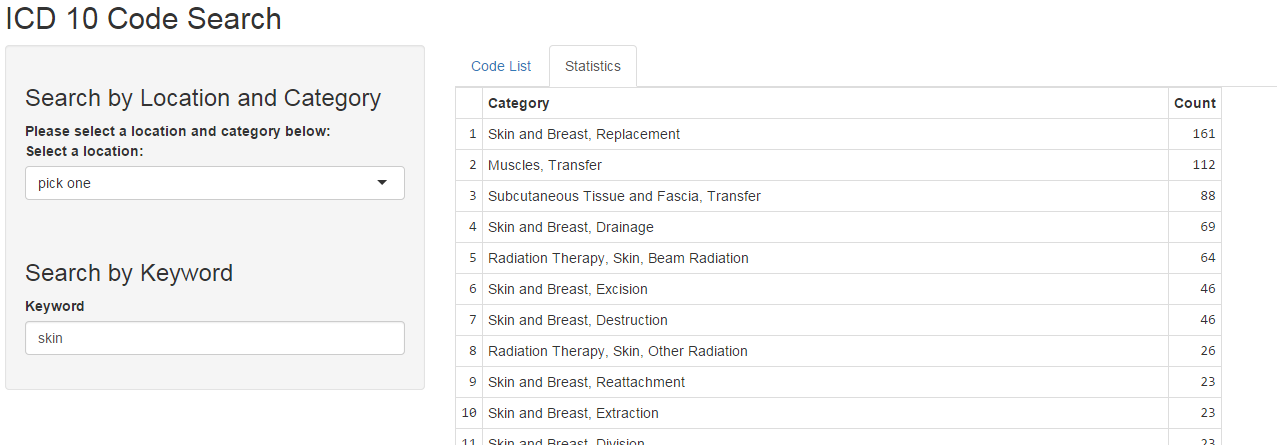
*Keyword Search*



The default results set that is displayed to the user is a list of codes and long descriptions that meet the search criteria. They are displayed in order according to their code. While it is the case that codes with the same parent will all be grouped together in the Code List, it does not display any indication of which Category each code belongs to. This can be easily updated in a further iteration of this application.

The second results set is displayed on a separate tab labeled Statistics. Currently, there is only one statistic that is presented to the user. When a user selects this tab, it will display, in decreasing order, a count of how many codes fall under each category for the given results set. This was done in order to help the user identify which category may be the most applicable to their search and thereby help them narrow the filters used so that a smaller results set can be obtained.

*Statistics Tab*



# IMPEDIMENTS

This project laid a good deal of foundation for enhancements to a basic search application for the ICD-10 Coding System. By using the graph database, there is a lot of potential for creating complex relationships without impairing the performance of the application. Because the majority of the data was contained within a single string on each node (Long Description), it was difficult to identify properties and relationships that can be extracted easily. One of the additional filter criteria that I would have like to include is a subset of the location filter. With the long descriptions of the codes, there are very often references to more specific locations. While this level of detail can be searched on in the keyword search feature, I think it would have added value to the dynamic filter as well.

The code itself also reveals information and relationships; however, the information it reveals is necessarily tied to the header which it falls under. Because there are many headers in the dataset, it did not seem appropriate to identify relationships by hand that only occur under a single header. There may be a more detailed mapping of the code to its meanings, but it was not included in this dataset or the supplementary documentation that came with it.

# EXTENSION

I propose that an extension to this project could include developing a recommendation list based on a user’s previous searches and selections. Nodes could be ranked by how often they appear in the results set and how often the user actually selects that code as the one they were looking for. The recommendation list could be prepopulated with the highest ranked codes, and would update dynamically based on any search criteria the user enters.

# CITED LITERATURE:

Rahmathulla, G., Deen, H. G., Dokken, J. A., Pirris, S. M., Pichelmann, M. A., Nottmeier, E. W., & ... Wharen Jr, R. E. (2014). Implementation and impact of ICD-10 (Part II). Surgical Neurology International, 5(4), S192-S198. doi:10.4103/2152-7806.137182

CDC/National Center for Health Statistics. 2013. International Classification of Diseases, (ICD-10-CM/PCS) Transition. Retrieved from: <http://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm>

Conn, Joseph. (2013). Expensive. Confusing. Time consuming. *Modern Healthcare*. Retrieved from: <http://www.modernhealthcare.com/article/20131026/MAGAZINE/310269932>

AAPC. (2014). What is Medical Coding?. Retrieved from: <https://www.aapc.com/medical-coding/medical-coding.aspx>

World Health Organization. (2014). International Classification of Diseases (ICD). Retrieved from: <http://www.who.int/classifications/icd/en/>

Bowman, Sue. (2008). "Why ICD-10 Is Worth the Trouble." Journal of AHIMA 79, no.3 24-29.

2015 ICD-10 PCS and GEMs. (2015). *2015 PCS Long and Abbreviated Titles* [Data File]. Retrieved from: http://www.cms.gov/Medicare/Coding/ICD10/2015-ICD-10-PCS-and-GEMs.html