Predicting Escalation of Musculoskeletal Care

Jacob Candela | Joeli McKee | Julia Quiñones | Chris Webb
Disclaimer

This is a project performed by M.S. students in Applied Economics. The Applied Project course is the capstone course for the degree of M.S. in Economics. The course is a three-month consulting project that takes the student from the proposal stage through the final presentation and final report.

The students are “hired” as associates in a fictitious consulting company called Applied Economics Research Group (AER Group). The consulting company is fictitious, and is only mentioned to enhance the realism of the project.

To further enhance the realism of the projects, we sometimes use hypothetical client requests for proposals with actual client names. **However, the client requests may not be real.** Please contact the Director of the Applied Master’s program at Florida State University to find out if a client request is real or not.

Although every effort has been made to verify the accuracy of the student projects, Florida State University makes no claims, promises, or guarantees about the accuracy, completeness, or adequacy of the projects, and disclaims liability for errors and omissions of content, statements and conclusions of the projects. In addition, note that these projects are the work of students, and do not necessarily reflect the views of the Department of Economics or Florida State University.

**Please do not quote, cite or distribute this project without written consent by the Director of the Applied Master’s program at Florida State University.**
Executive Summary

Background:
Integrated musculoskeletal care (IMC), a firm based in Tallahassee, is dedicated to reducing the incidence, prevalence, and costs associated with musculoskeletal disorders. Through a combination of claims data analysis, training, clinical care and prevention methods, IMC works with self-insured employers and healthcare organizations to improve clinical outcomes and reduce health care costs. IMC has tasked AERG with analyzing claims data to explain the patient diagnoses and procedure types that predict the likelihood of musculoskeletal escalation.

Data:
IMC provided AERG with a panel data set that contained over one million observations, describing 32,880 Medicare patients and their healthcare visits over the course of four years. Specifically, the data set contained information about patients’ diagnoses, the year of the first visit for each diagnosis, duration of treatment, the “pathway” of procedures for each diagnosis, and the total cost associated with each diagnosis. AERG was informed all patients are 65 years of age and older.

We observed that each patient has at least one diagnosis with most having between 1 and 150 diagnoses that range across all ICD-10 chapters. Having multiple conditions and/or disorders in the same patient is referred to as comorbidity. The overlap of medical conditions is very common, especially for older patients. Patients with musculoskeletal disorders are often diagnosed with other conditions such as diabetes, mental health disorders, and cardiovascular disease. AERG believes that having diagnoses in addition to a musculoskeletal diagnosis significantly impacts the likelihood a patient will have a musculoskeletal escalation.

Methods:
IMC defined musculoskeletal escalation as three separate categories: Surgery without MRI, MRI without surgery, and Surgery with MRI. AERG found that the number of patients that had a Surgery with MRI within the data was insufficient. Because of this we felt that we could not accurately predict if patients will have a Surgery with MRI. AERG also found that many patients had multiple escalations with different escalation categories. To account for these patients, AERG created Multiple Escalation as an additional category of escalation.

AERG considered several methods to predict patient escalation category including traditional econometric models as well as machine learning models. Our prediction methods were determined to be more superior to a naïve method – where patients’ number of diagnoses was the sole predictor in the likelihood of escalation.
Results:

After considering both econometric and machine learning methods, we determined that we cannot predict which category of escalation a patient will have but instead can best predict if a patient will have an escalation. AERG found that the Two-Class Boosted Decision Tree was the best model for predicting patient escalation.

<table>
<thead>
<tr>
<th>Prediction Method and Percent of Escalation Category Correctly Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
</tr>
<tr>
<td>Escalation</td>
</tr>
</tbody>
</table>

Recommendation:

AERG recommends IMC use a Two-Class Boosted Decision Tree to predict patient escalation. We can extrapolate these findings in order to quantify healthcare savings. AERG observed 19,307 patients who did not have an escalation with a median cost of $15,423. We also observed 13,573 patients who had an escalation with a median cost of $33,246. Using the Two-Class Boosted Decision Tree model we can identify 77 percent of the patients that escalate and save $290,073,041 in patient costs, approximately 18% of total patient costs.
# Table of Contents

Executive Summary ................................................................. 3  
Table of Contents ........................................................................ 5  
Introduction ................................................................................ 7  
  Musculoskeletal Disorders ......................................................... 7  
  Integrated Musculoskeletal Care .................................................. 8  
  AERG’s Goal ............................................................................. 11  
Data Description .......................................................................... 12  
Data Preparation .......................................................................... 16  
  Aggregating to the Patient Level ................................................. 16  
New Variable Creation ................................................................. 17  
  ICD-10 ................................................................................... 17  
  ICD-10 Chapters 1-21 ............................................................. 19  
  Wscore .................................................................................. 22  
  Procedure Categories ............................................................... 24  
  Year Indicator Variables ........................................................... 27  
Musculoskeletal Escalation ........................................................... 31  
Data Cleaning .............................................................................. 34  
  Censored Data ......................................................................... 34  
  Left-Censored data ................................................................... 35  
  Right-Censored data ................................................................. 35  
Methodology ................................................................................. 37  
Variables Used in Analysis ............................................................ 37  
  Model Inputs ............................................................................ 37  
  Model Outputs ......................................................................... 38  
Hypotheses .................................................................................. 38  
Model Selection ............................................................................ 40  
  Predicting versus Explaining ...................................................... 40  
  Predictive Modeling .................................................................. 40  
  Explanatory Modeling ............................................................... 41  
  Choosing Our Models ............................................................... 41  
  Econometric Models ................................................................. 42
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning</td>
<td>47</td>
</tr>
<tr>
<td>Results</td>
<td>53</td>
</tr>
<tr>
<td>Model Comparison</td>
<td>53</td>
</tr>
<tr>
<td>Econometric Regressions</td>
<td>54</td>
</tr>
<tr>
<td>Superior Model Determination</td>
<td>60</td>
</tr>
<tr>
<td>Variable Importance</td>
<td>60</td>
</tr>
<tr>
<td>Recommendation</td>
<td>62</td>
</tr>
<tr>
<td>Cost Savings IMC calculation</td>
<td>62</td>
</tr>
<tr>
<td>Works Cited</td>
<td>63</td>
</tr>
<tr>
<td>APPENDIX 1</td>
<td>65</td>
</tr>
<tr>
<td>APPENDIX 2</td>
<td>68</td>
</tr>
<tr>
<td>APPENDIX 3</td>
<td>77</td>
</tr>
<tr>
<td>APPENDIX 4</td>
<td>82</td>
</tr>
<tr>
<td>APPENDIX 5</td>
<td>87</td>
</tr>
</tbody>
</table>
**Introduction**

**Musculoskeletal Disorders**

Musculoskeletal disorders are injuries and diseases that affect the musculoskeletal system, including the skeleton, muscles, tendons, ligaments, and joints. According to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), there are over 450 unique musculoskeletal disorders. The most common of these disorders include sprains, fractures, tendonitis, various forms of arthritis, and osteoporosis.

Musculoskeletal disorders are extremely common in the United States and are among the leading causes of physical disability and pain. According to the United States Bone and Joint Initiative, one in two Americans aged 18 and older are living with a musculoskeletal disorder. This number jumps to three in four Americans when we consider people aged 65 and older. With an aging population, the United States is projected to experience an increase in these numbers.

In 2011, musculoskeletal disorders amounted to $213 billion in direct and indirect costs, which is roughly 1.5% of the United States Gross Domestic Product (GDP). Patients with musculoskeletal disorders are often diagnosed with other conditions such as diabetes, cardiovascular disease, and mental health disorders. In 2011, treatment costs for patients with both musculoskeletal and other conditions reached $796 billion, approximately 5% of the U.S. GDP.
Integrated Musculoskeletal Care

Integrated Musculoskeletal Care (IMC) has revolutionized the orthopedic industry through the application of response-based assessments and stringent care protocols that reduce the frequency, cost, and incidence of musculoskeletal disorders. IMC’s Outcome Accountable™ Care system provides patients with the most appropriate pathway of care for musculoskeletal disorders. In doing so, IMC is able to prevent patients from undergoing unnecessary, costly treatments and procedures in order to provide optimal patient outcomes.

Musculoskeletal disorders are often very difficult to diagnose. When patients report musculoskeletal pain to their doctors, the pain can stem from multiple locations and, in turn, be difficult to describe. Patients are often incorrectly diagnosed and administered inappropriate pathways of care. IMC specializes in accurately and consistently diagnosing musculoskeletal disorders, in order to counteract the 40% of musculoskeletal disorders that go misdiagnosed in the U.S. every year. Through the Outcome Accountable™ Care system, IMC is able to accurately diagnose musculoskeletal disorders and create specific treatment plans for patients based on their conditions. IMC’s treatment plans are developed with the intention of minimalizing the use of addictive pain medications or invasive procedures, such as injections and surgeries.

IMC works with self-insured employers to reduce costs associated with musculoskeletal disorders. Through their Outcome Accountable™ Care system, IMC guarantees a 25% savings in health care costs for self-insured employers. Not only is IMC able to reduce health care costs for employers, but they are also able to improve the quality of life and productivity of employees.
In addition to working with self-insured employers, IMC has expanded their analytics department to work with Accountable Care Organizations (ACOs). ACOs are groups of health care providers, including doctors and hospitals, who voluntarily coordinate high-quality care for patients. Recently they have begun assisting Medicare patients to help with coordinating care. Medicare is the federal health insurance program that provides coverage for people ages 65 and older and for certain younger people with disabilities. ACOs allow Medicare patients to access a network of coordinated providers who communicate across disciplines to identify the best care for each patient.

ACOs were created to correct for the lack of coordination of care in the U.S. health care industry. Consider the automotive repair industry as an example of efficient coordination. For example, imagine that your car will not start. You wouldn’t call a driver, who would then contact a battery mechanic who would then contact an alternator specialist who would then contact a starter consultant and transmission expert. You would probably take your car to the mechanic who has the tools and expertise to diagnose anything from a dead battery to engine failure.

ACOs are similar in the sense that they coordinate different components of care for patients—primary care, specialists, hospitals, home health care, etc. Basically, ACOs create a “one stop shop” where patients can obtain a diagnosis and review potential treatment plans. By streamlining this process of receiving a diagnosis and a treatment plan, ACOs improve care, avoid unnecessary tests and procedures, and reduce costs for patients.

An important distinction must be made here; ACOs save their clients’ money by coordinating care, which is only the front end of the transaction.
IMC saves their clients' money on the outcomes side, or the back end of the transaction. Given ACOs’ original purpose of saving costs by improving coordination, they are now looking for other areas of potential improvement. IMC can assist with this, as they specialize in preventative measures to cut health care costs in the long run.
**AERG’s Goal**

At AERG, we believe that our analysis will enable IMC to leverage its Accountable Care solutions to assist in reducing escalation of care and costs. By identifying diagnoses and procedures that predict the likelihood of musculoskeletal escalation, AERG can assist IMC in creating better coordination of care for their clients and continuing to reduce unnecessary treatments and procedures.

AERG will utilize both predictive and explanatory models to help IMC identify patients who are likely to escalate based on diagnoses and procedure types. With the ability to correctly predict patient escalation, IMC will be able to focus their attention on certain individuals who are more prone to escalation and high costs.

In particular, AERG plans to help IMC identify patients 65 and older who are likely to escalate. The number of people aged 65 and older is projected to grow from 15% of the population to 21% by the year 2040, making it increasingly important to identify and address health risks that could lead to costly procedures (The Burden of Musculoskeletal Diseases in the United States, 2015).

If the United States can bring quality, comprehensive healthcare to more people earlier in their lifetimes, it can cut down on the social and fiscal strains brought about by chronic diseases. The implications of improving coordination and efficiency on a national level are beyond the scope of our project, although they provide a useful illustration of the positive externalities that IMC is creating for its clients. Our goal is to aid IMC in the identification of patients who are likely to face increasing costs. IMC can then focus their efforts towards saving expenses by reducing unnecessary procedures.
Data Description

AERG was given a dataset from IMC that is comprised of Medicare patients. The data set was received as a comma-separated value (.csv) file that contains 1,500,000 rows and seven columns. Each row is a single observation, or “case”, and is associated with a unique diagnosis for a particular patient. The data set contains 32,880 unique patients.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Column Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMID</td>
<td>1</td>
<td>Patient identification code. De-identified, alphanumeric with 21 characters</td>
</tr>
<tr>
<td>CSID</td>
<td>2</td>
<td>Case Identification, also referred to as “case.” Patient identification code combined with ICD diagnosis code. Our data set contains 1,500,000 unique cases.</td>
</tr>
<tr>
<td>FVCS</td>
<td>3</td>
<td>First date of diagnosis. De-identified. Year and month only.</td>
</tr>
<tr>
<td>DUR</td>
<td>4</td>
<td>Days of treatment from the first visit date.</td>
</tr>
<tr>
<td>PXimage</td>
<td>5</td>
<td>String of procedures and services. Comma-separated.</td>
</tr>
<tr>
<td>ICD10</td>
<td>6</td>
<td>Diagnosis code used to identify diagnosis/illness. Contains both ICD-9 and ICD-10 codes. Our data set contains 23,263 unique diagnosis codes.</td>
</tr>
<tr>
<td>TCcs</td>
<td>7</td>
<td>Total cost for each case.</td>
</tr>
</tbody>
</table>

EMID

EMID stands for “Encrypted Member Identification” and is a de-identified code that represents each individual patient within the data. Patient identification is encrypted in order to comply with HIPAA regulations. Each EMID is a 21-character code that consists of letters and numbers. For instance, an EMID could look like: 18908BCDEIJGHFJHU5444. AERG identified 32,880 unique patients within the data set.
CSID

CSID stands for “Case Identification” and is a concatenation of the EMID and a particular diagnosis. Each CSID, also referred to as a “case” is associated with an individual patient. All patients have at least one case, while most patients have about 30 cases.

Figure: Histogram of Number of Diagnoses

FVCS

First Visit Case is a date variable that represents the date of the first visit for each case. In order to comply with HIPAA regulations, FVCS contains the year and month of the first visit but does not contain the day. The data spanned four years from October 2014 to November 2018. The majority of patients had their first visit in 2015.

DUR

Duration is the number of days of treatment for each case. Duration was calculated by IMC as the last date of treatment minus the first date of treatment. The values for duration range from 0 to 1141 days. DUR has a mean of 105 days and a median of 0 days. AERG assumes that
cases associated with a duration of 0 represent instances in which a problem is diagnosed and treated within the same visit.

PXimage

PXimage was given to us in the form of a single string that represents a chronological list of medical services and procedures for each case, also referred to as a “procedure pathway.” AERG identified over 300,000 unique procedure pathways. Each PXimage string contains CPT codes, which are 5-digit numbers separated by a comma. Most PXimages contain 1 CPT, while the maximum number of codes within a PXimage for any case is 75 codes. There are no repeated CPT codes within the PXimage string. Although a patient may have received more than one of the same procedure for a particular diagnosis, it is not demonstrated in the data. See the following section for more on PXimages.

ICD10

ICD10 represents “International Classification of Diseases” diagnosis that each patient has for a given case. ICD10 contains both ICD-9 and ICD-10 codes. The data set contains 23,263 unique ICD codes. See the following section for more on ICD codes.

TCcs

TCcs is the total cost of treatment for each case. The values for TCcs range from $.051 to $2,565,548.79. TCcs has a mean of $1087.45 and a median of $161.23.
Below is a simulated data set that is meant to demonstrate the structure of the data we were given. Our data set was unique at the case level, meaning that each row of data was a unique CSID. The data contains repeat observations at the EMID (patient) level, as shown in the second column of the sample set below. Each observation of PXimage contains a certain number of CPT codes, which begin with an apostrophe and end with a comma.

Table: Simulated Data Set

<table>
<thead>
<tr>
<th>obs</th>
<th>Emily</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M299008</td>
</tr>
<tr>
<td>2</td>
<td>17190</td>
</tr>
<tr>
<td>3</td>
<td>f99</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>4809</td>
</tr>
<tr>
<td>7</td>
<td>988</td>
</tr>
<tr>
<td>8</td>
<td>9999</td>
</tr>
<tr>
<td>9</td>
<td>17190</td>
</tr>
<tr>
<td>10</td>
<td>f99</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>13</td>
<td>4809</td>
</tr>
<tr>
<td>14</td>
<td>988</td>
</tr>
<tr>
<td>15</td>
<td>9999</td>
</tr>
<tr>
<td>16</td>
<td>17190</td>
</tr>
<tr>
<td>17</td>
<td>f99</td>
</tr>
<tr>
<td>18</td>
<td>f7629</td>
</tr>
<tr>
<td>19</td>
<td>c48902</td>
</tr>
<tr>
<td>20</td>
<td>d90</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>4809</td>
</tr>
<tr>
<td>24</td>
<td>988</td>
</tr>
<tr>
<td>25</td>
<td>9999</td>
</tr>
<tr>
<td>26</td>
<td>17190</td>
</tr>
<tr>
<td>27</td>
<td>f99</td>
</tr>
<tr>
<td>28</td>
<td>f76298</td>
</tr>
<tr>
<td>29</td>
<td>T1212</td>
</tr>
<tr>
<td>30</td>
<td>g8903</td>
</tr>
<tr>
<td>31</td>
<td>f99</td>
</tr>
<tr>
<td>32</td>
<td>f76298</td>
</tr>
<tr>
<td>33</td>
<td>h6690</td>
</tr>
</tbody>
</table>
Data Preparation

Aggregating to the Patient Level

Several factors contributed to AERG’s decision to aggregate IMC’s data to the patient level for analysis. Ideally, AERG would be able to construct a timeline of events for each of IMC’s 32,880 patients. However, with the absence of repeated procedure codes and the presence of overlapping treatment for different diagnoses, we were unable to construct a theoretically sound timeline of treatment. Instead we created a rich database of patient characteristics by aggregating the data we were given. AERG utilized sophisticated data mapping and data mining techniques to flag diagnoses and procedures of interest. This process is detailed in the following sections.
New Variable Creation

ICD-10

Created by the World Health Organization (WHO), the International Classification of Diseases (ICD) codes are a uniform vocabulary of diagnostic codes that describe causes of injury, illness and death. ICD codes are revised periodically by the National Center for Health Statistics (NCHS) to account for clinical modifications. The number following “ICD” represents the revision of the code that is in use.

In 2015, the Department of Health and Human Services instructed all healthcare providers to transition from ICD-9 to ICD-10 codes in order to improve the quality of diagnosis documentation. ICD-9, which contained 13,000 codes, had reached its capacity as a coding system and could not account for new diagnoses. ICD-10 contains 68,000 codes and increases diagnosis specificity by including additional subcategories and sub classifications with regards to the cause, manifestation, location, severity and type of injury or disease.
<table>
<thead>
<tr>
<th>ICD-10 improves upon ICD-9 coding conventions by:</th>
<th>ICD-10 improves upon ICD-9 functionality by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Providing 5 times the amount of unique diagnostic codes as provided by ICD-9</td>
<td>• Allowing for direct comparisons of mortality and morbidity rates on the state, national, and international level</td>
</tr>
<tr>
<td>• Cementing the first character of a diagnostic code as an easily-identifiable alpha character</td>
<td>• Allowing for more descriptive measurement of an individual’s illness over time, which in turn will provide a more accurate measurement of the quality of care that patient receives</td>
</tr>
<tr>
<td>• Allowing for diagnostic codes to span 3-7 characters, as opposed to 3-5 characters</td>
<td>• Indicating on which side of the body the illness/injury occurs</td>
</tr>
<tr>
<td></td>
<td>• Indicating severity</td>
</tr>
<tr>
<td></td>
<td>• Indicating complexity of illness/injury condition</td>
</tr>
<tr>
<td></td>
<td>• Updating terminology and concepts to adhere with current medical practice and understanding</td>
</tr>
<tr>
<td></td>
<td>• Providing higher quality information about injuries with external causes like trauma or poisoning (cause, severity, location on the body, code extensions for details not accounted for otherwise)</td>
</tr>
<tr>
<td></td>
<td>• Distinguishing between complications that occur during and after procedures</td>
</tr>
</tbody>
</table>

The World Health Organization transitioned ICD-9 over to ICD-10 codes in 2015. Our data set contains observations that span from 2014 to 2018, which includes the transition from ICD-9 to ICD-10. We used ICD Code-recognition software to confirm that our data contained both ICD-9 and ICD-10 codes. Due to the differences in the levels of specificity, it is not possible to directly compare ICD-9 codes and ICD-10 codes.

AERG then utilized the General Equivalence Mapping (GEM) created by the Centers for Medicare and Medicaid Services to convert the ICD-9 Codes to ICD-10 codes. We created a new variable called ICD-10 that contains only ICD-10 codes.
ICD-10 Chapters 1-21

After creating the ICD-10 variable, we were about to identify 18,464 unique ICD-10 codes in our data. AERG decided to group the ICD-10 codes by chapters, as created by the WHO. ICD-10 codes are divided into twenty-one chapters and sorted alphabetically. Each chapter refers to one of the body’s systems and is specified by a range which is displayed in the first three digits of the code.

AERG created count variables for each ICD-10 chapter. Listed directly below is a table of the ICD-10 chapters (along with a short description) and their assigned variable names. For example, the ICD-10 codes between A00-B99 are accounted in the variable CH1.
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>RANGE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A00-B99</td>
<td>Certain infections and parasitic diseases</td>
</tr>
<tr>
<td>2</td>
<td>C0-D49</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>3</td>
<td>D50-D89</td>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</td>
</tr>
<tr>
<td>4</td>
<td>E00-E89</td>
<td>Endocrine, nutritional, and metabolic diseases</td>
</tr>
<tr>
<td>5</td>
<td>F01-F99</td>
<td>Mental, Behavioral and neurodevelopmental disorders</td>
</tr>
<tr>
<td>6</td>
<td>G00-G99</td>
<td>Diseases of the nervous system</td>
</tr>
<tr>
<td>7</td>
<td>H00-H59</td>
<td>Diseases of the eye and adnexa</td>
</tr>
<tr>
<td>8</td>
<td>H60-H95</td>
<td>Diseases of the ear and mastoid process</td>
</tr>
<tr>
<td>9</td>
<td>I00-I99</td>
<td>Diseases of the circulatory system</td>
</tr>
<tr>
<td>10</td>
<td>J00-J99</td>
<td>Diseases of the respiratory system</td>
</tr>
<tr>
<td>11</td>
<td>K00-K95</td>
<td>Diseases of the digestive system</td>
</tr>
<tr>
<td>12</td>
<td>L00-L99</td>
<td>Diseases of the skin and subcutaneous tissue</td>
</tr>
<tr>
<td>13</td>
<td>M00-M99</td>
<td>Diseases of the musculoskeletal system and connective tissue</td>
</tr>
<tr>
<td>14</td>
<td>N00-N99</td>
<td>Diseases of the genitourinary system</td>
</tr>
<tr>
<td>15</td>
<td>O00-O99</td>
<td>Pregnancy, childbirth, and puerperium</td>
</tr>
<tr>
<td>16</td>
<td>P00-P96</td>
<td>Certain conditions originating in the perinatal period</td>
</tr>
<tr>
<td>17</td>
<td>Q00-Q99</td>
<td>Congenital malformations, deformations, and chromosomal abnormalities</td>
</tr>
<tr>
<td>18</td>
<td>R00-R99</td>
<td>Symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified</td>
</tr>
<tr>
<td>19</td>
<td>S00-T88</td>
<td>Injury, poisoning, and certain other consequences of external causes</td>
</tr>
<tr>
<td>20</td>
<td>V00-Y99</td>
<td>External causes of morbidity</td>
</tr>
<tr>
<td>21</td>
<td>Z00-Z99</td>
<td>Factors influencing health status and contact health services</td>
</tr>
</tbody>
</table>
The data set included observations within every ICD-10 chapter, excluding chapter 15. The five most frequently observed ICD-10 chapters are listed in the table below.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
<th>Examples</th>
<th>Number of Cases</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified</td>
<td>Shortness of breath, cardiac murmur</td>
<td>237,234</td>
<td>30,625</td>
</tr>
<tr>
<td>13</td>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>Muscle tear, fractured radius</td>
<td>206,872</td>
<td>27,528</td>
</tr>
<tr>
<td>9</td>
<td>Diseases of the circulatory system</td>
<td>Hypertension, pulmonary heart disease</td>
<td>178,319</td>
<td>31,364</td>
</tr>
<tr>
<td>21</td>
<td>Factors influencing health status and contact health services</td>
<td>Problems related to lifestyle, postpartum care</td>
<td>158,015</td>
<td>31,183</td>
</tr>
<tr>
<td>4</td>
<td>Endocrine, nutritional, and metabolic diseases</td>
<td>Diabetes, premature menopause</td>
<td>117,435</td>
<td>28,271</td>
</tr>
</tbody>
</table>

When inspecting the diagnosis categories at the case level, musculoskeletal diagnoses do not comprise the majority of observations. When evaluating the data at the patient level however, 27,528 out of the 32,880 distinct patients (87%) had at least one musculoskeletal diagnosis.

Because the data set is heavily comprised of observations of diagnoses other than musculoskeletal, AERG believes that the presence of other diagnoses is related to the likelihood of having a musculoskeletal disorder diagnosis and escalation.
Considering that most of the patients in the data set have multiple diagnoses, AERG felt it was necessary to get a better understanding of the relationship between musculoskeletal diagnoses and other diagnoses.

After conducting a brief literature review, AERG observed there is often a correlation between musculoskeletal disorders and other disorders classified by ICD-10 codes. The most prevalent diagnoses that occur in patients with musculoskeletal disorders are cardiovascular diseases, diabetes, and respiratory diseases. Often these conditions lead to prolonged inactivity or the development of musculoskeletal pain, resulting in workplace absences and increased medical attention. Diabetes and cardiovascular disease are oftentimes chronic conditions which lead to musculoskeletal disorders such as carpal tunnel syndrome and arthritis. Several studies have shown that depression and anxiety can lead to a significant increase in musculoskeletal pain, especially if fatigue is present in the patient.

The presence of two or more conditions within a patient is defined as comorbidity. We decided to include a variable that would contain a numerical measure for comorbidity. To create a measurement for comorbidity, AERG chose to include an already established and widely used comorbidity index, the Charlson Comorbidity Index. The Charlson Comorbidity Index (CCI) categorizes comorbidities of patients based on the International Classification of Diseases (ICD) diagnosis codes.

The index includes thirteen conditions, assigns weights (from 1 to 6) to each condition, and then sums the weights to produce a single comorbidity score for each patient.
The CCI was originally developed to predict the likelihood of a one-year mortality rate for a patient. A patient with a CCI score of zero indicates that no comorbidities were found. The higher the score, the more likely the predicted outcome will result in mortality or higher resource use.

Including the Charlson Comorbidity Index is imperative, due to the high correlation between several diagnoses included in the CCI and the development of musculoskeletal disorders. Considering the chronic nature of the conditions categorized by the CCI, this index provides us a way to easily identify patients who are at very high risk of escalating. Mainly, the weighted comorbidity score helps to create a profile for the patients in our dataset and allows us to analyze their overall health status.

Below is a table that exhibits the conditions included in the Charlson Comorbidity Index and the assigned weights.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial Infarction</td>
<td>1</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>1</td>
</tr>
<tr>
<td>Peripheral Vascular Disease</td>
<td>1</td>
</tr>
<tr>
<td>Cerebrovascular Disease</td>
<td>1</td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease</td>
<td>1</td>
</tr>
<tr>
<td>Dementia</td>
<td>2</td>
</tr>
<tr>
<td>Paraplegia and Hemiplegia</td>
<td>1</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1</td>
</tr>
<tr>
<td>Diabetes with Complications</td>
<td>2</td>
</tr>
<tr>
<td>Renal Disease</td>
<td>2</td>
</tr>
<tr>
<td>Mild Liver Disease</td>
<td>1</td>
</tr>
<tr>
<td>Moderate or Severe Liver Disease</td>
<td>3</td>
</tr>
<tr>
<td>Peptic Ulcers</td>
<td>1</td>
</tr>
<tr>
<td>Rheumatic Disease</td>
<td>1</td>
</tr>
<tr>
<td>Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS)</td>
<td>6</td>
</tr>
<tr>
<td>Cancer</td>
<td>2</td>
</tr>
</tbody>
</table>
Procedure Categories

The data set contains a string variable PXimage that is comprised of all the procedures a patient underwent for each diagnosis. PXimage is made up of a string of procedure codes, commonly known as CPT codes (example procedure “pathway” for a single diagnosis: 85027, '36415, '82248, '81001, and'80053).

Current Procedural Terminology (CPT) codes are derived from the American Medical Association and are used in the medical billing process. CPT codes provide the insurance payer with the procedures for which the healthcare provider wishes to be reimbursed. These codes give a detailed description of evaluations, tests, surgeries, and any other medical procedures that may be performed on a patient. When combined with ICD-10 diagnosis codes, CPT codes help to create a sufficient picture of the medical process for the insurance payer. “The patient arrived with these symptoms (as represented by the ICD-10 code) and we performed these procedures (represented by the CPT code).”

The CPT code set is exceptionally large and includes codes for tens of thousands of medical procedures. The codebook is broken down into three main categories, labeled as category 1, 2, and 3. Category 1 is the most relevant when dealing with large healthcare datasets because categories two and three are not as frequently used in a day to day setting.

Similar to the way that ICD-10 codes are divided into chapters, Category 1 CPT codes are divided into six sections. For each of IMC’s 1,500,000 individual procedure pathways, we identified the number of procedures that fell into each category.
We then created six new variables (one for each procedure category) to store this information at the case level. Below is a list of the six subsets along with their designated ranges:

- **Evaluation and Management (99201 – 99499):**
  - The Evaluation and Management category primarily consist of primary care visits in which the patient is examined and given a diagnosis.

- **Anesthesia (00100 – 01999; 99201 – 99410):**
  - The Anesthesia category includes anesthesia services that are provided for procedure on specific body areas such as the head, neck thorax, spine, upper leg, or elbow.

- **Surgery (10021 – 69990):**
  - The Surgery category is the largest section within the CPT code set and is organized by the part of the body the surgery is performed on.

- **Radiology (70010 – 79999):**
  - The radiology section is comprised of codes for a CT scan, MRI, MRA, and nuclear medicine.

- **Pathology and Laboratory (80047 – 89398):**
  - Pathology being the study of disease, pathologists interpret biopsies (an examination of tissue removed from a living body to discover the presence, cause, or extent of a disease, an examination of tissue removed from a living body to discover the presence, cause, or extent of a disease), monitor laboratory testing, while aiding in the interpretations of the laboratory tests. This billing code range is organized into clinical
pathology and anatomic pathology. Anatomic pathology refers to the examination of bone, joints tendons, ligaments, and muscle.

- **Medicine (90281-99199; 99500-99607):**
  - The Medicine category covers a wide range of procedures, an example being a vaccine or dialysis.

- **Non-MSK Surgery (Procedure code range in Appendix 2):**
  - The Non-MSK Surgery category was defined by IMC as all surgical procedures which do not belong to our musculoskeletal surgery escalation category. It is important to take these procedures into account when determining a patient’s likelihood to escalate to musculoskeletal procedures, considering that surgical complications can result in chronic disorders in the long term.
Considering that our patients entered the data set at different points in time over the course of four years, 2014 to 2018, we added a measure to account for the issue of endogeneity. Specifically, we wanted to account for the amount of time each patient was observed in the data. We created five binary variables that indicate the year that the patient first entered the data, one for 2014, 2015, 2016, 2017, and 2018.

**Year Indicator Variables**

- Considering that our patients entered the data set at different points in time over the course of four years, 2014 to 2018, we added a measure to account for the issue of endogeneity. Specifically, we wanted to account for the amount of time each patient was observed in the data. We created five binary variables that indicate the year that the patient first entered the data, one for 2014, 2015, 2016, 2017, and 2018.
Total Duration

Using our year indicator variables, we identified the time between when the patient first entered the data and the last date that they were observed. We then created a continuous variable that identifies the total duration the patient was observed, in years. We included our duration measure in the form of a dummy variable. This variable classifies a patient as being observed in the data for less than one year, one year, two years, three years, or four years.

<table>
<thead>
<tr>
<th>Time Observed</th>
<th>Percentage of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than One Year</td>
<td>4%</td>
</tr>
<tr>
<td>One Year</td>
<td>11%</td>
</tr>
<tr>
<td>Two Years</td>
<td>14%</td>
</tr>
<tr>
<td>Three Years</td>
<td>28%</td>
</tr>
<tr>
<td>Four Years</td>
<td>43%</td>
</tr>
</tbody>
</table>

Figure: Time Observed in the Data
Total Cost

We created a new total cost variable that is a sum of all of the patient’s cases. Total Costs for each patient ranges from $23.17 to $2,920,390.37 with a median of $22,396.84.

Figure: Total Cost per Patient (Logged)
We also found that patients who were observed in the data longer had higher costs. As you can see from the figure below, the costs are the largest among patients that have been in the data for four years.

**Figure: Total Costs by Duration (In Years)**
Musculoskeletal Escalation

IMC defined musculoskeletal escalation as three distinct categories: surgery without MRI, MRI without surgery, and surgery with MRI. Before making any predictions about which of the categories a patient will escalate to, AERG first wanted to find the number of patients within each escalation category. A few issues arose:

While the categories of musculoskeletal escalation were defined for us, they were not included as their own variables within the data set. Thus, AERG created three binary variables: MSK_Surg (surgery without MRI), MSK_MRI (MRI without surgery), and MSK_SM (surgery with MRI). To create each variable, we identified the ranges of CPT procedure codes for each escalation category. The procedure codes for musculoskeletal surgery range from 20100 to 29999. If a patient had a PXimage that contained a CPT procedure code within the musculoskeletal surgery range, MSK_Surg would be coded as a 1.

The procedure codes for musculoskeletal MRI range from 01922 to 78815. If a patient had a PXimage that contained a procedure code within the musculoskeletal MRI range, MSK_MRI would be coded as a 1. If a patient had a PXimage that contained procedure codes for both a surgery and an MRI, then MSK_SM was coded as a 1.

Out of the 32,880 unique patients within the data set, only 77 patients (0.02% of the patient population) had a surgery with an MRI. Due to such a small number of patients belonging to this category, models predicting the probability of a musculoskeletal surgery with MRI will be unstable. Thus, AERG chose to omit models with the dependent variable as surgery with MRI.
During our analysis AERG also found that there were patients in the data set who had escalations in more than one of the defined escalation categories. For example, a patient could have a shoulder pain diagnosis and have a PXimage that include a surgery with an MRI. That same patient could also have a diagnosis for Osteoarthritis and have a PXimage that includes an MRI without surgery. AERG determined that an additional category of escalation needs to be included in our analysis that accounts for if patients have had multiple musculoskeletal escalations. AERG created the binary variable MSK_MULT_ESC that is coded as a 1 when a patient has multiple cases that were classified as different escalation categories. Given the escalation categories discussed above, the graphics below describe some basic attributes of each category.
<table>
<thead>
<tr>
<th>No MSK Escalation</th>
<th>MSK Surgery without MRI</th>
<th>MSK MRI without Surgery</th>
<th>Multiple Escalations</th>
</tr>
</thead>
<tbody>
<tr>
<td>19,307 patients</td>
<td>6,606 patients</td>
<td>2,625 patients</td>
<td>4,292 patients</td>
</tr>
<tr>
<td>58.7% of patients</td>
<td>20.1% of patients</td>
<td>7.8% of patients</td>
<td>13.2% of patients</td>
</tr>
<tr>
<td>Total cost range: $23.17 - $2,228,354.49 total cost per patient</td>
<td>Total cost range: $393.21 - $2,920,390.37 total cost per patient</td>
<td>Total cost range: $672.15 - $1,661,196.84 total cost per patient</td>
<td>Total cost range: $1,675.6 - 1,059,866.4 total cost per patient</td>
</tr>
<tr>
<td>Median total cost for each patient is $15,422.95</td>
<td>Median total cost for each patient is $29,028.43</td>
<td>Median total cost for each patient is $27,810.71</td>
<td>Median total cost for each patient is $43,625.63</td>
</tr>
<tr>
<td>Median of the total duration patients were observed in the data: 3 years</td>
<td>Median of the total duration patients were observed in the data: 4 years</td>
<td>Median of the total duration patients were observed in the data: 4 years</td>
<td>Median of the total duration patients were observed in the data: 3 years</td>
</tr>
<tr>
<td>32 diagnoses (median)</td>
<td>48 diagnoses (median)</td>
<td>50 diagnoses (median)</td>
<td>65 diagnoses (median)</td>
</tr>
</tbody>
</table>
Our newly defined escalation categories show a more complete picture of the patients who escalate. There is a clear correlation with escalation and increased costs. In fact, multiple escalations show an even increased cost. This shows not only what our escalating patient looks like, but also the opportunity for savings. The picture of our escalating patient solidifies the need for further analysis to reduce escalation and save Medicare costs.

**Data Cleaning**

**Censored Data**

As previously discussed, our data set contains 32,880 unique patients with an observation time of roughly four years. However, the length of our duration variable for each patient can vary dramatically. For example, one patient can enter the data set in 2015 and exit in 2016, providing AERG with a limited amount of knowledge on the patient’s history. Another example of this problem is when we observe patients who enter in 2015 and exit in November 2018. In this case, we are assuming that a diagnosis has been resolved, even though we have incomplete knowledge of the patient’s treatment for this diagnosis. Both cases cause a censoring issue for our model interpretations.

When dealing with healthcare data, censoring is a prevalent issue. Left-censored and right-censored data frequently disrupts the accuracy of a prediction model. The basic principle behind censored data is that you are unable to know if a patient had prior treatments before entering the data set. For example, if someone enters the data with an office visit and concludes with a surgery we cannot determine if that patient had a surgery or any other treatments before the initial office visit. The same complication arises when dealing with right censored data. With certain data the study ends before you can observe a predicted outcome or event occurring. For
example, if a study ends after four years those individuals who have not escalated in the 4 years could be considered censored. We do not know if the individual hasn’t escalated because they were cured or because the study had ended.

![Figure: Censored Data Example](Image)

**Left-Censored data**

Left-censored data is especially a frequent issue when dealing with healthcare data. We do not know if someone had any prior treatment based on the four-year set of data we were given. There was no prior history attached to the patients before the year 2014. For the left-censored data issue, we could not do anything to provide a suitable solution. We must assume an individual could have been previously treated for any number of complications.

**Right-Censored data**

We did attempt to address the right-censoring issue in a number of ways, but it continued to be a difficult problem to overcome. Our solution to correct for right-censored data was to eliminate observations that occurred within the last three months of the data set. The assumption
was made that these individual patients did not complete treatment or were removed from the data for irregular reasons.

**Further Attempts**

We also tried to address the issue of endogeneity in our dataset by subsetting the data to patients who had a duration of 3.5 years or greater. This was used as an attempt to create a static duration variable for the set of patients. However, we dismissed this option due to the steep reduction of our number of patients from 32,880 to approximately 14,000, essentially eliminating over 50% of the patients from our data set. Simply eliminating a substantial amount of our observations was deemed an unjustifiable approach to solving the censored data issue.
Methodology

Variables Used in Analysis

Model Inputs

- Total Diagnosis Count
- Total Procedure Count
  - 6 measures defined in the “Procedure Codes” subsection
- ICD-10 Chapter Counts
  - 21 measures defined in the “Diagnosis Codes” subsection
- Charleston Comorbidity Weighted Score
- Patient Cost
  - Total Cost for All Cases
- Time Components
  - Year of First Case
  - Total Duration
- 5 indicator variables defined in the “Data Structure” section
- Total Duration Observed
- 5 indicator variables defined in the “Data Structure” section
Model Outputs

- We used several different variations of musculoskeletal escalation as our dependent variable.

1. **Predicting Escalation**: General definition of escalation that does not differentiate between the different categories of escalation: Surgery without MRI, MRI without surgery, and Multiple Escalation. In this case, the dependent variable could take one of two values: “No Escalation” or “Escalation.”

2. **Predicting Surgery without MRI**: Escalation takes one of two values: “Surgery without MRI” or “No Escalation.”

3. **Predicting MRI without surgery**: Escalation takes one of two values: “MRI without Surgery” or “No Escalation.”

4. **Predicting Multiple Escalation**: Escalation takes one of two values: “Multiple Escalation” or “No Escalation.”

5. **Predicting Escalation Category**: Escalation takes one of three four values: “Multiple Escalation” or "Surgery without MRI”, “MRI without surgery”, or “No Escalation.” Models with this output predict which escalation category a patient will have based on their characteristics such as diagnoses and procedure types.

Hypotheses

After conducting a topical literature review and exploratory data analysis, AERG proposed several hypotheses about the relationships between patient characteristics and certain escalation outcomes. A scatterplot matrix is included below to illustrate the reasoning behind some of these hypotheses.
As demonstrated by the above graphic, the strongest statistical relationship between two input variables involves number of diagnoses and total number of procedures. Number of diagnoses and number of procedures are strongly positively correlated, with no outliers and very little variation. The relationship between number of diagnoses and number of non-musculoskeletal surgeries is also positively correlated, although to a lesser extent. While we identify a slightly positive correlation between number of diagnoses and weighted CCI score, there is too much noise to draw a conclusion about their relationship.
We also find that certain escalation categories tend to form clusters within these relationships. For instance, we can see that patients who do not escalate (marked by a green x) tend to fall in lower values of both number of diagnoses and number of procedures. When evaluating the relationship between number of diagnoses and total duration, we find that non-escalating patients constitute the majority of patients who were observed for a year or less while patients with multiple escalations tend to be in the data for three or more years.

**Model Selection**

**Predicting versus Explaining**

The framework of our methodology begins with our valuation of explanatory power versus predictive power in our models. In order to demonstrate the discrepancy between these two approaches, we provide a brief overview of explanatory and predictive models.

**Predictive Modeling**

Predictive modeling utilizes large sets of input variables to produce predicted values of a single output variable. Because the primary goal of a predictive model is to produce values with the highest measure of fit possible, the weights given to the input variables are not necessarily interpretable. However, predictive analytics is useful for producing consistent predictions which rely on many input variables that take on many possible values. In particular, predictive modeling software like Machine Learning utilizes data mining to identify patterns within qualitative input variables. For instance, a video streaming website may utilize information about its visitors and the videos they watch to predict and suggest more videos to them. Machine Learning would be useful in this case because it can identify recurring key words in video titles.
Unfortunately, because predictive modeling lacks explanatory power, it is not useful for determining the root causes of correlation among variables. We will utilize Machine Learning to assess the potential predictive power of our input variables.

**Explanatory Modeling**

Explanatory modeling allows us to examine the relationship between specific independent and dependent variables. By utilizing explanatory modeling, we can identify which input variables are given the most weight when determining the output variables. We can substantiate our theoretical claims about the relationship between two variables through econometric analysis. Econometric analysis relies on quantitative criterion, which must be met in order to make a claim about the relationship between an input and out variable. For instance, we can determine a particular variable’s significance by evaluating its p-value and the magnitude of its coefficients. We may also use quantitative criterion to evaluate the effectiveness of the model as a whole, taking into account the significance and magnitude of all input variables. While evaluating the model as a whole, we will inspect the p-value of the global F-test, the percent of predicted values which are concordant and discordant, and its ROC curve value. We will use Statistical Analytical Software (SAS) to execute our econometric analysis and evaluate the explanatory power of our model.

**Choosing Our Models**

The two prevalent methods for classification are econometric modeling and machine learning. Econometric models are used to interpret the predictor variables for a given outcome. These help to explain the changes that researchers can see when looking at a problem. Machine Learning methods can do both. The algorithm used in machine learning are the differentiator in
this respect. Choosing a logistic regression algorithm in a machine learning tool gives the researcher explanatory power. Using a different algorithm such as a neural network or a decision tree provide fewer explanatory results. Machine learning models are typically better at making predictions.

**Econometric Models**

Several econometric models were considered including a basic ordinary least squares model (OLS) which is commonly used as a generalized linear modelling technique. The OLS model is designed to be used with one independent (explanatory) variable or several independent variables. The model also assumes a continuous dependent variable and attempts to create a linear relationship between the explanatory values and the dependent variable to create a line of best fit. In other words, it will create a straight line through the data points improving the efficiency of the model. It is important to note that this line of best fit is based on the conditional means (simply the mean, calculated after a set of prior conditions has happened) of the explanatory variables.

An ordinary least squares approach allows you to see the effects of certain outlier values in your data set in approximation to the means. It also allows you to check a linearity model assumption if that is what your data set seems to present. Another time OLS is used is when we want to see if there is much disparity amongst variances.

However, the OLS model is often times not used because of its limitations. It is restricted to a linear relationship which often times is not effective.
For example, when dealing with age and income the relationship is not linear. The model can be extremely responsive to outliers as well depending on the makeup of the explanatory variables.

After considering the ordinary least squares regression we determined it would not be a reasonable approach for several reasons. The dependent variables for our models are categorical variables due to the defined escalation grouping. The OLS model often assumes a normal distribution which is often not the case in healthcare data. An ordinary least squares approach also does not allow us to control for perfect collinearity. For example, Body Mass Measurements (BMI) and obesity status would be too correlated to use in a model. For these reasons, we decided the OLS model would not be able to lead us to an unbiased prediction.

We also attempted running a hurdle model during our model selection process. Hurdle models are actually the amalgamation of two different models. First the data is run through a
binary logistic regression, which predicts if a person will fall into the “0” category (non-escalation) or the “1” category (escalation). If a person falls into the non-escalation category, they will not be used to estimate the second regression. The second regression is a count data model, which is used to predict the degree to which the person will experience escalation.

AERG attempted to correct for overrepresented zeroes by running several hurdle models. These models proved to be unstable because they were unable to meet the convergence criterion. In essence, the model was unable to find a significant ‘peak’ within the data, which would identify a patient as escalating or non-escalating. Because of this, the logistic component of the model was unreliable. This makes the count model component unreliable as well, which invalidates the model as a whole.

**Logistic Regression**

*Binary Logistic Regression*

Unlike OLS, which predicts a continuous dependent value from continuous and categorical inputs, binary models are used to address the probability that an observation falls into one of two mutually exclusive classifications. Logistic regression is one example of such binary models. Consider a situation in which a patient undergoes blood testing to determine if they are diabetic. In this situation, the practitioner must use information about the patient to diagnose them as diabetic (diabetic=1) or not (diabetic=0). Here our outcome variable is categorical rather than continuous. However, our inputs can still take both continuous and categorical values.

If a doctor was assigned the task of diagnosing this patient, they would most likely assess the patient’s A1C levels. A1C measures the percentage of the patient’s hemoglobin that contains
glucose and is a continuous measure. AERG are tasked with predicting whether a patient will experience an escalation of musculoskeletal care, which requires much more information than a single blood test can provide. Because of this, we will be utilizing both continuous and categorical variables to estimate the likelihood of a patient escalating.

A logistic regression is part of the generalized linear model family which allows for a response variable that is non-normally distributed. The basic idea of the generalized model is that a link function changes the scale of the response variable to match the output of a linear model. In the logistic model, the link function is the logit function.

\[ \text{Logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_1 x + \epsilon \]

This logit link function constrains the fitted values to always fall between 0 and 1, creating a binary outcome. The figure shown below demonstrates how the distribution of a model is transformed by the logistic link function. In a normal distribution (shown left), most values will cluster around the mean. In a logistic distribution (shown right), all values will display as either 0 or 1, with the mean being somewhat arbitrary.
Binomial logistic regression was used for prediction with the first four model outputs: *Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without Surgery*, and *Predicting Multiple Escalation*.

*Multinomial Logistic Regression*

Multinomial logistic regression is used when the dependent variable can take more than two values. An extension of binary logistic regression, multinomial logistic regression is used to predict categorical placement. Multinomial logistic regression assumes that the dependent variable categories are independent. Furthermore, this assumption asserts that the choice of one category is not related to the choice of another category. To test this assumption, we used the Hausman-McFadden test.

Multinomial logistic regression was used to predict the fifth model output: *Predicting Escalation Category*. The multinomial model uses patients’ diagnoses and procedure types to predict which escalation category they will be categorized as: “No Escalation”, “MRI without Surgery”, “Surgery without MRI”, or “Multiple Escalation.”
Machine learning is a branch of artificial intelligence that automates complex mathematical algorithms to big data. Machine learning is able to analyze data and make decisions with limited human intervention. There are two types of machine learning techniques available to any statistician or economist: supervised and unsupervised learning. Supervised learning is used for classification and regression modeling. Unsupervised learning is used in clustering and association learning problems. Machine learning models generally have more predictive power than econometric models, however machine learning predictors frequently cannot be used for explanatory analysis.
Unsupervised Learning

Unsupervised learning is a form of artificial intelligence that uses algorithms to find patterns and identify groups within the data. Often these patterns are not able to be seen with the human eye. Unsupervised algorithms perform on their own, without human guidance, and present underlying distributions and structures within the data.

For example, music streaming services use unsupervised machine learning to make song recommendations to listeners based on their listening history and song preferences. Unsupervised machine learning algorithms are able to detect specific instruments, sounds, and tempos to recommend similar songs, bands, and even genres of music. For example, the instruments or tempos used in some country songs may be similar to those in another genre of music, like gospel. Country music listeners may be recommended to certain gospel songs or bands due to the patterns that unsupervised machine learning algorithms were able to identify, patterns that humans are unable to detect.

Considering that unsupervised machine learning algorithms produce previously unknown outputs, AERG ultimately felt that unsupervised machine learning was not appropriate to predict defined the categories of musculoskeletal escalation. Instead, AERG chose to utilize supervised machine learning.

Supervised Learning

The majority of machine learning uses supervised learning. Supervised learning algorithms creates predictions based on labeled data. The algorithms then utilize patterns and associations to apply labels to new data. Unlike unsupervised machine learning, the supervised learning environment does not make suggestions that are beyond human comprehension, unless
specified to do so. It all depends on the type of algorithm that is chosen within the supervised learning arena. With supervised machine learning, a data scientist guides the algorithms to produce labeled output.

Supervised learning can be split into two categories, classification and regression. Classification models are used to predict qualitative outputs. These models use the independent variables to determine the class to which the dependent variable belongs. For instance, classification mode below uses animal characteristics and determines if they are a cat or a dog. Regression models, on the other hand, predict quantitative outputs. These models, like the one pictured below, uses independent variables to predict numerical values, i.e. house prices.

AERG chose to use two types of supervised machine learning models that use classification algorithms, Two-Class Boosted Decision Tree and Multi-Class Decision Tree as well as Two-Class Neural Networks and Multi-Class Neural Networks.
Tree Based Models

Two-Class Boosted Decision Tree and Multi-Class Random Forest are both Tree based models. Tree based machine learning models are used frequently for their superior prediction power. When compared with other machine learning models, these models are an attractive approach due to their simplicity.

Decision trees are useful when the dependent variables are categorical and can be divided into two or more unique populations. In other words, a decision tree splits the data into separate categories based on the values of the independent variables.

The Two-Class Boosted Decision Tree models were used to predict our first four model outputs: Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without Surgery, and Predicting Multiple Escalation.

AERG used the Multi-Class Random Forest model to predict our fifth model output: Predicting Escalation Category. A random forest when used for the purpose of classification, assembles several decision trees when the model is being trained. However, instead of simply gathering the average predictions of each individual decision tree a random forest builds the trees based on randomized sampling of the data. Also, when the trees are split into “Nodes” the random forest randomizes these subsets. This method helps correct for the decision trees’ tendency to over fit the training set by implementing what is known as Bootstrapping or Bagging (for further explanation of “bagging” please see Appendix 1). The main advantage of a random forest compared to a decision tree is the model’s ability to correct for overfitting.
Neural Networks

A neural network is set of interconnected layers that is designed to resemble the human brain. Neural Networks are typically used in unsupervised machine learning for clustering but can also be used in supervised machine learning for classification. The inputs are the initial layer where the data is fed into the algorithm. Input layers are connected to an internal layer known as the interconnected layer. This is affectionately known as the “black box” to most economists. This layer is also the place where the unseen connections within the data are made. There is no explanatory power within this layer. The output layer has the output nodes which are the classifiers.
Although deep neural networks have been shown effective with complex tasks, we choose to not use a deep neural network. Instead we chose to use a Two-Class Neural Network and Multi-Class Neural Network as these models are conducted through supervised machine learning and are useful when predicting classes of outcomes.

The Two-Class Neural Network model was used to predict our first four model outputs: Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without Surgery, and Predicting Multiple Escalation. AERG used the Multi-Class Neural Network model to predict our fifth model output: Predicting Escalation Category.
Results

Model Comparison

AERG ran all the models listed above on each of our model outputs: Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without Surgery, Predicting Multiple Escalation, and Predicting Escalation Category. We looked at three measures to compare the models: sensitivity, specificity, and accuracy.

Sensitivity, also known as the true positive rate, is the percentage of the positive class that is correctly predicted by the model. For example, in our models predicting escalation vs. no escalation, the sensitivity is the percentage of escalations correctly predicted.

\[
\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}
\]

Specificity, also known as the true negative rate, is the percentage of non-events correctly predicted by the model. In our models predicting escalation vs. no escalation, the specificity is the percentage of no escalations correctly predicted.

\[
\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}
\]

Accuracy is derived from the sensitivity and specificity of the model and measures the overall accuracy. It is a measure of the correctly predicted events and non-events out of the total number of predictions.
Econometric Regressions

Binomial Logistic Regression

As you can see our binary logistic regressions did not accurately predict any of the model outputs: Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without surgery, and Predicting Multiple Escalation.

<p>| Figure: Sensitivity Analysis for Binary Logistic Regressions |
| Binary Logistic Regression |</p>
<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Model Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>70.57%</td>
</tr>
<tr>
<td>Specificity</td>
<td>89.18%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>79.88%</td>
</tr>
<tr>
<td>Naïve Accuracy</td>
<td>64.35%</td>
</tr>
</tbody>
</table>

Our sensitivity analysis reveals that our chosen logistic model performs worse than our naïve model when predicting Surgery without MRI and MRI without Surgery. While we were able to improve our accuracy in terms of predicting general escalation and multiple escalation, our predictive power increased only marginally (24.13% and 30.36%, respectively). We determined that we cannot accurately predict if a patient will have a musculoskeletal escalation using binary logistic regression. While binary logistic regression is
unable to accurately predict patient escalation, we still feel it is important to discuss the important variables.

**Important Variables**

In order to compare the magnitude of our regression coefficients, we ran several binomial regressions with different groupings of input variables. These groupings include:

1) Patient-level counts of diagnoses, broken out by ICD-10 category
2) Patient-level counts of non-escalating procedures, broken out into 6 categories
3) Indicator variables for total duration of patient treatment (using less than one year of treatment as a reference group)
4) All variables listed above, in addition to number of diagnoses, total procedures, and weighted CCI score

Due to the nature of our problem and binomial regressions, we ran four sets of regressions: each variable grouping was used to estimate **Predicting Escalation, Predicting Surgery without MRI, Predicting MRI without Surgery, and Predicting Multiple Escalation.**

**Concerning ICD-10 Categories**

When we compare the effects of diagnosis category on any level of escalation, we find that musculoskeletal diagnoses have the most consistent effect on our patient population. For each additional musculoskeletal diagnosis, a patient is approximately one and a half times more likely to escalate. Out of our four variable groupings, our ICD-10 variables produced the most
consistent predictions for escalation. Our diagnosis categories gave us 88% concordant predictions, meaning that 88% of the patients who escalated were correctly predicted.

**Concerning Procedure Categories**

Comparing the effects of different procedure categories on escalation, we find that anesthesia has the greatest effect on whether a patient will escalate to any level, but particularly for patients who escalate to MRI without surgery. For each additional anesthesia treatment, a patient becomes approximately three times more likely to escalate to MRI without surgery and one and a half times more likely to escalate to surgery without MRI. This contradicts our initial expectation, which was that anesthesia would be more highly correlated with surgical procedures than MRIs. We believe that this may be due to the inclusion of injections within our surgery indicator. Because injections are common and minimally invasive (not requiring anesthesia), the importance of anesthesia in predicting more invasive surgeries is discounted within our data.

**Concerning Total Duration Indicators**

Using patients who were treated for less than a year as a reference group, each subsequent group (one year, two years, three years, and four years) was more likely to escalate than the group before. A patient moving from less than one year of observation to at least one year of observation experiences approximately the same increase in risk of escalation as a patient moving from three years of observation to four years. The increased risk associated with moving from one to two years and from two to three years is smaller in magnitude, compared to the two groups discussed prior.
Multinomial Logistic Regression

Our multinomial logistic model was approached in the same way that our binary logistic model was executed. Each patient was assigned either a 0 or a 1 for each of our four escalation categories. We then compared each patient’s predicted escalation to their actual escalation status, but the sensitivity and specificity measures discussed earlier no longer apply. Because sensitivity and specificity rely on the true positive and true negative rate of predictions (and our multinomial regression no longer predicts a binary), we instead use an average measure of accuracy.

Figure: Sensitivity Analysis for Multinomial Logistic Regression

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>68.72%</td>
<td>40.33%</td>
<td>27.89%</td>
<td>61.51%</td>
</tr>
<tr>
<td>Naïve Accuracy</td>
<td>64.35%</td>
<td>50%</td>
<td>50%</td>
<td>52.83%</td>
</tr>
</tbody>
</table>

Our sensitivity analysis for the multinomial regression reveals that the prediction of separate escalation categories is severely hindered for both the naïve and chosen model by including the dependent variable as a multinomial. While our chosen model improves on the naïve model for every escalation category, our chosen multinomial model has far less predictive power than our chosen logistic models. We determined that we cannot predict which of the escalation categories a patient will be identified using multinomial logistic regression.
Machine Learning

Figure: Random Forest vs. Multi-Class Neural Network

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Random Forest</th>
<th>Multi-Class Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Output</td>
<td>5. Predicting Escalation Category</td>
<td>5. Predicting Escalation Category</td>
</tr>
<tr>
<td>Accuracy</td>
<td>65%</td>
<td>69%</td>
</tr>
</tbody>
</table>

We used the Random Forest and Multi-Class Neural Network to test which escalation category a patient will belong to, our categories being no escalation, surgery without MRI, MRI without surgery, and multiple escalation. Because these models produced relatively low accuracy, we determined that we cannot predict the escalation category.
**Figure: Sensitivity Analysis for Binary Logistic Regressions**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td>77%</td>
<td>66%</td>
<td>49%</td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td>78%</td>
<td>71%</td>
<td>78%</td>
<td>57%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td>82%</td>
<td>80%</td>
<td>82%</td>
<td>89%</td>
<td></td>
</tr>
</tbody>
</table>

**Figure: Sensitivity Analysis for Neural Networks**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td>72%</td>
<td>51%</td>
<td>51%</td>
<td>59%</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td>82%</td>
<td>71%</td>
<td>61%</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td>82%</td>
<td>80%</td>
<td>84%</td>
<td>88%</td>
<td></td>
</tr>
</tbody>
</table>
**Superior Model Determination**

While both the Two-Class Boosted Decision Tree and the Two-Class Neural Network using the fourth model output: **Predicting Multiple Escalation** resulted in the highest accuracy, they resulted in the relatively low sensitivity. Sensitivity measures how well we accurately predict patient escalation. Ultimately, we determined that sensitivity is the most important measure to consider when choosing the superior model.

Out of all the models, the Two-Class Boosted Decision Tree that uses the first model output: **Predicting Escalation** has the highest sensitivity of 77%. Therefore, we believe that we can most accurately predict whether a patient will have a musculoskeletal escalation where escalation does not differentiate between the different categories: Surgery without MRI, MRI without surgery, and Multiple Escalation.

**Variable Importance**

While the Two-Class Boosted Decision Tree with the first model output: **Predicting Escalation** does not produce interpretable coefficients, we are able to derive the importance of each variable on the predictive power of the model.

Not surprisingly, the number of musculoskeletal disorder diagnoses has the largest impact on the predictive power of the model. Also not surprising, the number of total procedures and the number of diagnoses has a significant impact on the predictive power of the model. This is because both of these measures are, in effect, a proxy for the amount of times we observe each patient. For instance, the more diagnoses a patient has, the more “cases” we observed for that patient. Additionally, the more procedures a patient has, the more likely it is that one of those
procedures will be classified as escalation. Also unsurprisingly, each patient’s total cost had a significant impact in predicting escalation.

Number of Chapter 18 diagnoses was also significant in predicting escalation for our patient population. Chapter 18 includes “symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified.” AERG believes that the number of Chapter 18 diagnoses is important in predicting escalation because it indicates that a patient could be experiencing complications with their health. For instance, let us say a patient undergoes bloodwork because they have experienced chronic headaches and/or fatigue. This could be a sign of an impending diabetes diagnosis. Since we know that musculoskeletal disorders are linked with other chronic disorders, including diabetes, chapter 18 could be used to determine if someone is at risk to develop a musculoskeletal disorder, and by extension, a musculoskeletal escalation. This is confirmed by the fact that pathology and laboratory procedures are also significant in prediction of escalation.
Recommendation

Cost Savings IMC calculation

AERG recommends IMC use a Two-Class Boosted Decision Tree to predict patient escalation. We can extrapolate these findings in order to quantify healthcare savings.

AERG observed 19,307 patients who did not have an escalation and had a median cost of $15,423. We also observed 13,573 patients who had an escalation and had a median cost of $33,246. With the Two-Class Boosted Decision Tree model we can identify 77 percent of the patients that escalate and save $290,073,041 in patient costs, approximately 18% of the total patient costs.
Works Cited

“About Pathology and Laboratory Medicine.” Hospital for Special Surgery, Hospital for Special Surgery, 2019, www.hss.edu/pathology-about.asp.


APPENDIX 1

Suggestions for Future Analysis

**Age and Gender:** The data AERG received was limited in some ways in regard to patient characteristics. Factors such as age and gender were not included in the data set which restricted our ability to categorize patients and control for certain independent variables in the models. Based on previous literature, we determined that a gender variable would not have definitively played a significant role in the econometric predictive models, yet it would have been a variable worth considering. Also, gender might have a positive effect on certain machine learning predictive methods.

Without a variable for age it was difficult for AERG to establish basic information on an individual patient. Age would have certainly had a significant role in a predictive model for the established escalation categories. Most likely, we would have seen notable marginal returns from our age variable due to the strong correlation between age and health status. Also, age would have allowed AERG to control for several diagnosis chapters. For example, as age increases the risk of acquiring a circulatory disease (CH9) increases drastically.

**Medicare Data:** The data was limited to only patients who qualified for Medicare which allows us to observe individuals who are at least 65 years of age or individuals who are at least 18 years of age and disabled. This hinders AERG’s ability to properly look at patients who represent the general population. We are unable to compare escalation patterns amongst providers as well due to the fixed nature of the variable. It would have been extremely compelling if AERG was able to observe the effects of certain providers on escalating to a surgery, MRI, or multiple escalations.
**Current Procedural Terminology (CPT Codes):** The procedure pathways which contained our CPT billing codes was found to be coded as a string variable. In other words, we could not distinguish a singular CPT code from our PXimage (Procedure Pathway) variable. The variable being a string makes it extremely difficult to find patterns involving the CPT codes and our escalation categorical variables.

It was also troublesome to identify structure amongst the CPT billing codes due to the fact we did not have any repeating codes. For example, if a patient was billed for a Physical Therapy session, we are unable to discern if that individual went to one PT meeting or several. This restricts our predictive ability a large amount in regard to a patient’s pathway.

**Comorbidity Index:** The data did not have an established comorbidity index which can be a significant predictor in healthcare data. AERG used the Charlson comorbidity index to replace the missing variable but this index has certain characteristics that were not ideal for the data set. For example, the Charlson comorbidity index is measured using a one-year mortality for its weighted wscore. This is not ideal because we are simply looking to see if a patient escalates to a surgery, MRI, or multiple escalation; we are not looking at a patient’s risk of mortality.

**Duration:** The patients throughout our entire data set were not like cohorts. Meaning, our group of individuals did not have similar statistical factors. The main issue came from the fluctuations in duration amongst our patients. As previously discussed, one patient may be observed for one year while another may be observed for several years. Ideally, we would like to observe only patients with similar durations. This would assist in solving the matter of our left and right censored data.
Unsupervised Learning

PX image: Our procedure pathway was given as a non-repeating pathway for each diagnosis. This means that a patient with a physical therapy procedure was only counted once. The procedures were not repeating. This limits the analysis that can be done because the possibility of certain procedures preceding escalations to MRI or surgeries could not be accounted for. For example, if a patient escalated to MRI after a physical therapy procedure, it was unknown if the physical therapy procedure was in one visit or spread across multiple physical therapy visits prior to receiving the MRI.

One solution to this would be using unsupervised learning to determine the clusters of procedure pathways. This would allow for further analysis with a new set of interesting parameters. Given the scope of this endeavor AERG would recommend this as a way to identify possible pathway correlations with escalation. Reviewing the characters as part of a whole like Deep Learning analysis could be a solution for the PX image variable.

Bagging: A tool that can be used to overcome an unpredictable data set in machine learning is bagging (also known as “Bootstrapping). This is a technique that combines the algorithms of multiple machine learning methods in order to improve the predictions of a specific model. In other words, “Bootstrap Aggregation is a general procedure that can be used to reduce the variance for those algorithm that have high variance” (Brownslee, 2016). AERG was unable to produce an appropriate bagging method but it is recommended that IMC considers this approach to improve upon the decision tree model provided.
APPENDIX 2

Data Preparation

Load in necessary packages for data preparation

```r
library(dplyr)
library(stringr)
library(sqldf)
library(icd)
library(epiDisplay)
```

Read in original data set given by IMC

```r
C <- read.csv("C:\Users\AERG\AERG2019v2.csv")
```

Load in ICD9 General Equivalency Mapping (GEM) and convert all ICD9 codes to ICD10

```r
D <- read.csv("C:\Users\AERG\ICD9conversion.csv")
#Delete duplicates of ICD9 codes
D <- D %>% distinct(ICD9, .keep_all = TRUE)
#Convert ICD9 to ICD 10
C1 <- sqldf("SELECT C.X, C.ICD10, C.EMID, C.CSID, C.FVCS, C.DUR, C.NPI, C.PXimage,
D.ICD10 AS ICD9
  FROM C
  LEFT JOIN D
  on C.ICD10=D.ICD9")
```

Remove invalid observations of ICD codes and ICD9 codes.

```r
#Report ICD10 when ICD9 is Null
C2 <- sqldf("SELECT *
```
CASE WHEN ICD9 IS NULL THEN ICD10
ELSE ICD9
END AS ICD
FROM C1
ORDER BY ICD")

#Drop invalid observations of ICD codes
C3<-subset(C2, ICD>'999')

#Drop original ICD10 variable and newly created ICD9 variable
C3<-select(C3,-c(ICD10,ICD9))

#Rename C3 as C
C<-C3

Create ICD10 chapter variables (1-21)

C1<- sqldf( "SELECT *, CASE WHEN ICD LIKE 'A%' OR ICD LIKE 'B%' THEN 1
WHEN ICD LIKE 'C%' OR ICD LIKE 'D0%' OR ICD LIKE 'D1%' OR ICD LIKE 'D2%' OR ICD LIKE 'D3%' OR ICD LIKE 'D4%' THEN 2
WHEN ICD LIKE 'D5%' OR ICD LIKE 'D6%' OR ICD LIKE 'D7%' OR ICD LIKE 'D8%' THEN 3
WHEN ICD LIKE 'E%' THEN 4
WHEN ICD LIKE 'F%' THEN 5
WHEN ICD LIKE 'G%' THEN 6
WHEN ICD LIKE 'H0%' OR ICD LIKE 'H1%' OR ICD LIKE 'H2%' OR ICD LIKE 'H3%' OR ICD LIKE 'H4%' OR ICD LIKE 'H5%' THEN 7
WHEN ICD LIKE 'H6%' OR ICD LIKE 'H7%' OR ICD LIKE 'H8%' OR ICD LIKE 'H9%' THEN 8
WHEN ICD LIKE 'I%' THEN 9
WHEN ICD LIKE 'J%' THEN 10
WHEN ICD LIKE 'K%' THEN 11
WHEN ICD LIKE 'L%' THEN 12
WHEN ICD LIKE 'M%' THEN 13
"
WHEN ICD LIKE 'N%' THEN 14
WHEN ICD LIKE 'O%' THEN 15
WHEN ICD LIKE 'P%' THEN 16
WHEN ICD LIKE 'Q%' THEN 17
WHEN ICD LIKE 'R%' THEN 18
WHEN ICD LIKE 'S%' OR ICD LIKE 'T%' THEN 19
WHEN ICD LIKE 'V%' OR ICD LIKE 'W%' OR ICD LIKE 'X%' OR ICD LIKE 'Y%' THEN 20
WHEN ICD LIKE 'Z%' THEN 21
END AS ICD_CAT
FROM C")

Load the comorbidity package and calculate the Charlson Comorbidity Index for each patient

```
library(comorbidity)
C2<-comorbidity(x = C1, id = "EMID", code = "ICD", score = "charlson")
```

Create Surgery Binary Variable

```
#Step 1: create a numeric range for surgery
surg_cpt <- "([2][0-9][1-9][0-9-9][0-9])"
Surg <- str_match(C$PXimage, surg_cpt)
s1 <- as.data.table(Surg)
s2 <- s1$V1
CS2 <- cbind(C,s2)
CS2$surg <- ifelse(!is.na(CS2$s2),1,0)
C<-CS2
```

Create MRI variable

```
#Create MRI CPT code list creation
MRI_cpt1 <- "([7][0][4][5-8][0])"
```
#Create single MRI variable in the CS2 data table

CS2$MRI <- ifelse(CS2$MRI1 == 1, 1, ifelse(CS2$MRI2 == 1, 1, ifelse(CS2$MRI3 == 1, 1, ifelse(CS2$MRI4 == 1, 1, ifelse(CS2$MRI5 == 1, 1, ifelse(CS2$MRI6 == 1, 1, ifelse(CS2$MRI7 == 1, 1, ifelse(CS2$MRI8 == 1, 1, ifelse(CS2$MRI9 == 1, 1, ifelse(CS2$MRI10 == 1, 1, ifelse(CS2$MRI11 == 1, 1, ifelse(CS2$MRI12 == 1, 1, ifelse(CS2$MRI13 == 1, 1, ifelse(CS2$MRI14 == 1, 1, ifelse(CS2$MRI15 == 1, 1, ifelse(CS2$MRI16 == 1, 1, ifelse(CS2$MRI17 == 1, 1, ifelse(CS2$MRI18 == 1, 1, ifelse(CS2$MRI19 == 1, 1, ifelse(CS2$MRI20 == 1, 1, ifelse(CS2$MRI21 == 1, 1, ifelse(CS2$MRI22 == 1, 1, ifelse(CS2$MRI23 == 1, 1, ifelse(CS2$MRI24 == 1, 1, ifelse(CS2$MRI25 == 1, 1, ifelse(CS2$MRI26 == 1, 1, ifelse(CS2$MRI27 == 1, 1, ifelse(CS2$MRI28 == 1, 1, 0)))))))))))))))))))))))

Create Radiology variable

#Create MRI CPT code list creation

rad_cpt1 <- "([7][0][0-3][1-9][0-5])"
rad_cpt2 <- "([7][0][0-3][1-9][7-9])"
rad_cpt3 <- "([7][0][4][5-9][0-9])"
rad_cpt4 <- "([7][0][5][5][4-9])"
rad_cpt5 <- "([7][1][0-2][0-9][0-9])"
rad_cpt6 <- "([7][1][5][5][5])"
rad_cpt7 <- "([7][2][0][0-9][0-9])"
rad_cpt8 <- "([7][2][1][0-3][0-9])"
rad_cpt9 <- "([7][2][1][7][0])"
rad_cpt9 <- "([7][2][1][9][0-4])"
rad_cpt10 <- "([7][2][2][0-9][0-9])"
rad_cpt11 <- "([7][3][0-1][0-9][0-9])"
rad_cpt12 <- "([7][3][2][0-1][0-9])"
rad_cpt13 <- "([7][3][2][2][5])"
rad_cpt14 <- "([7][3][0-6][0-9][0-9])"

rad_cpt15 <- "([7][3][7][0][0-9])"

rad_cpt16 <- "([7][3][7][2][5])"

rad_cpt17 <- "([7][4][0-1][0-7][0-9])"

rad_cpt18 <- "([7][4][0-1][0-7][0-9])"

rad_cpt19 <- "([7][4][0-1][8][4-9])"

rad_cpt20 <- "([7][4-6][0-3][0-6][0-9])"

rad_cpt21 <- "([7][4-6][0-3][8-9][0-9])"

rad_cpt22 <- "([7][6][5-9][0-9][0-9])"

rad_cpt23 <- "([7][7][0][0-1][0-9])"

rad_cpt24 <- "([7][7][0][2][1])"

rad_cpt25 <- "([7][7][0][4][6])"

rad_cpt26 <- "([7][7][0][7][1-7])"

rad_cpt27 <- "([7][7][0][8][1-3])"

rad_cpt28 <- "([7][7][0][8][5-9])"

rad_cpt29 <- "([7][7][2-7][0-9][0-9])"

rad_cpt30 <- "([7][8][0-2][0-9][0-9])"

rad_cpt31 <- "([7][8][4-9][0-9][0-9])"

rad_cpt32 <- "([7][9][0-9][0-9][0-9])"

#Create single radiology variable in the C data table

C$rad <- ifelse(C$rad1 == 1, 1, ifelse(C$rad2 == 1, 1, ifelse(C$rad3 == 1, 1, ifelse(C$rad4 == 1, 1, ifelse(C$rad5 == 1, 1, ifelse(C$rad6 == 1, 1, ifelse(C$rad7 == 1, 1, ifelse(C$rad8 == 1, 1, ifelse(C$rad9 == 1, 1, ifelse(C$rad10 == 1, 1, ifelse(C$rad11 == 1, 1, ifelse(C$rad12 == 1, 1, ifelse(C$rad13 == 1, 1, ifelse(C$rad14 == 1, 1, ifelse(C$rad15 == 1, 1, ifelse(C$rad16 == 1, 1, ifelse(C$rad17 == 1, 1, ifelse(C$rad18 == 1, 1, ifelse(C$rad19 == 1, 1, ifelse(C$rad20 == 1, 1, ifelse(C$rad21 == 1, 1, ifelse(C$rad22 == 1, 1, ifelse(C$rad23 == 1, 1, ifelse(C$rad24 == 1, 1, ifelse(C$rad25 == 1, 1, ifelse(C$rad26 == 1, 1, ifelse(C$rad27 == 1, 1, ifelse(C$rad28 == 1, 1, ifelse(C$rad29 == 1, 1, ifelse(C$rad30 == 1, 1, ifelse(C$rad31 == 1, 1, ifelse(C$rad32 == 1, 1, 0)))))))))))))))))))))))))))))))"
Create Anesthesiology variable

```r
#Create Anesthesiology CPT code list creation
anest1_cpt <- "([0][0-1][1-9][0-9][0-9])"
anest2_cpt <- "([9][9][1][0-4][0-9])"
```

Create Management Evaluation variable

```r
#Create MRI CPT code list creation
manage_cpt <- "([9][9][2-4][0-9][0-9])"
```

Create Pathology Laboratory variable

```r
#Create MRI CPT code list creation
pathlab_cpt <- "([8][0-9][0-9][0-9][0-9])"
medic_cpt1 <- "([9][0-8][0-9][0-9][0-9])"
medic_cpt2 <- "([9][9][0-1][0-9][0-9])"
medic_cpt3 <- "([9][9][5-6][0-9][0-9])"
```

Create Non-musculoskeletal Surgery variable

```r
non_msk_surg_cpt1 <- "([1][0-9][0-9][0-9][0-9][0-9])"
non_msk_surg_cpt2 <- "([3-6][0-9][0-9][0-9][0-9][0-9])"
```

Find the first date each patient entered the data

```r
# change FVCS to actual working date
date <- paste(C$FVCS, "-01", sep = "")
date <- as.Date(date)
C$FVCS <- date
```
# find start date
# order by EMID, date
C <- C[order(C$EMID, C$FVCS),]
dups <- duplicated(C$EMID)
fv <- subset(C, dups == FALSE, sel = c(EMID, FVCS))
C$Pat_Start <- fv$FVCS[match(C$EMID, fv$EMID)]

# remove duplicated EMID for reference list D1
dups <- duplicated(C$EMID)
C1 <- subset(C, dups == FALSE)

# create alt ID ref for using in subset routine
C1$Indx2 <- 1:32880

# match on Index to establish IDalt
C$IDalt <- C1$Indx2[match(C$CSID, C1$CSID)]

unique(C$Pat_Start)
start_date <- C %>%
  group_by(EMID, Pat_Start)

  st_date2 <- sqldf("select distinct EMID, count(*) as num_csid, Pat_Start from C group by EMID order by Pat_Start")

  st_date2$year <- str_sub(st_date2$Pat_Start, 1, 4)

  st_date2$year <- as.factor(st_date2$year)

  tab1(st_date2$num_csid, st_date2$year)

Find the last data the patient was observed in the data

# find end date
# order by EMID, date
C <- C[order(C$EMID, C$FVCS, decreasing = TRUE),]
dups <- duplicated(C$EMID)
fv <- subset(C, dups == FALSE, sel = c(EMID, FVCS))
C$Pat_fin <- fv$FVCS[match(C$EMID, fv$EMID)]
summary(C$Pat_fin)

# remove duplicated EMID for reference list D1
dups <- duplicated(C$EMID)
C1 <- subset(C, dups == FALSE)

# create alt ID ref for using in subset routine
C1$Indx2 <- 1:32880

# match on Index to establish IDalt
C$IDalt <- C1$Indx2[match(C$CSID, C1$CSID)]
unique(C$Pat_Start)
start_date <- C %>% group_by(EMID, Pat_fin)
m <- as.matrix(c(1:10))
m <- as.duration(round(diff(C$Pat_fin, C$Pat_Start, units = "weeks"))/52))
C$e_dur <- m

Aggregate the data to the patient level

C1 <- sqldf("SELECT DISTINCT EMID, count(*) as numcsid, sum(CH13) as numMSKdiag,
sum(MSK_Surg) as totsurg, sum(MSK_MRI) as totmri, sum(MSK_SM) as totsm, (Pat_fin -
Pat_Start)/365 as EMID_totdur, sum(CH1) as totCH1, sum(CH2) as totCH2, sum(CH3) as
totCH3, sum(CH4) as totCH4, sum(CH5) as totCH5, sum(CH6) as totCH6, sum(CH7) as
totCH7, sum(CH8) as totCH8, sum(CH9) as totCH9, sum(CH10) as totCH10, sum(CH11) as
totCH11, sum(CH12) as totCH12, sum(CH13) as totCH13, sum(CH14) as totCH14, sum(CH15) as
totCH15, sum(CH16) as totCH16, sum(CH17) as totCH17, sum(CH18) as totCH18,
sum(CH19) as totCH19, sum(CH20) as totCH20, sum(CH21) as totCH21, sum(non_msk_surg) as
tot_non_msk_surg, sum(rad) as Radiology, sum(anest) as Anesthesiology, sum(manage) as
Eval_Management, sum(pathlab) as Pathology_Lab, sum(medic) as Medicine, sum(num_CPT) as
totCPT, sum(TCcs) as totcost, Wscore FROM C GROUP BY EMID")
**APPENDIX 3**

Summary Statistics Broken Out by Escalation Category

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Median</th>
<th>Lower Quartile</th>
<th>Upper Quartile</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>num_diag</td>
<td>32880</td>
<td>45.6202859</td>
<td>29.2656309</td>
<td>40.0000000</td>
<td>24.0000000</td>
<td>61.0000000</td>
<td>1.0000000</td>
<td>299.0000000</td>
</tr>
<tr>
<td>totcpt</td>
<td>32880</td>
<td>105.4162713</td>
<td>73.1221906</td>
<td>91.0000000</td>
<td>52.0000000</td>
<td>141.0000000</td>
<td>0</td>
<td>740.0000000</td>
</tr>
<tr>
<td>wscore</td>
<td>32880</td>
<td>3.1715633</td>
<td>2.7826417</td>
<td>3.0000000</td>
<td>1.0000000</td>
<td>5.0000000</td>
<td>0</td>
<td>21.0000000</td>
</tr>
<tr>
<td>totcost</td>
<td>32880</td>
<td>49.609.56</td>
<td>80.524.25</td>
<td>22.396.84</td>
<td>868.73</td>
<td>5836.91</td>
<td>23.1700000</td>
<td>2920390.37</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>32880</td>
<td>7.3980535</td>
<td>5.8897643</td>
<td>6.0000000</td>
<td>3.0000000</td>
<td>11.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH1</td>
<td>32880</td>
<td>1.0579988</td>
<td>1.4119794</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>14.0000000</td>
</tr>
<tr>
<td>totCH2</td>
<td>32880</td>
<td>1.5636861</td>
<td>2.4297211</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>30.0000000</td>
</tr>
<tr>
<td>totCH3</td>
<td>32880</td>
<td>0.5875608</td>
<td>1.1509442</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>13.0000000</td>
</tr>
<tr>
<td>totCH4</td>
<td>32880</td>
<td>3.5716241</td>
<td>3.3879294</td>
<td>3.0000000</td>
<td>1.0000000</td>
<td>5.0000000</td>
<td>0</td>
<td>42.0000000</td>
</tr>
<tr>
<td>totCH5</td>
<td>32880</td>
<td>0.7308698</td>
<td>1.6109614</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>51.0000000</td>
</tr>
<tr>
<td>totCH6</td>
<td>32880</td>
<td>1.2260036</td>
<td>1.7573504</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totCH7</td>
<td>32880</td>
<td>2.3852190</td>
<td>2.9672360</td>
<td>1.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>28.0000000</td>
</tr>
<tr>
<td>totCH8</td>
<td>32880</td>
<td>0.5857056</td>
<td>1.2987708</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totCH9</td>
<td>32880</td>
<td>5.4233273</td>
<td>5.1728677</td>
<td>4.0000000</td>
<td>2.0000000</td>
<td>7.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH10</td>
<td>32880</td>
<td>2.3914842</td>
<td>3.2564619</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH11</td>
<td>32880</td>
<td>1.7360706</td>
<td>2.6700371</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>46.0000000</td>
</tr>
<tr>
<td>totCH12</td>
<td>32880</td>
<td>1.7732056</td>
<td>2.4389364</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>39.0000000</td>
</tr>
<tr>
<td>totCH13</td>
<td>32880</td>
<td>6.2917275</td>
<td>6.9858271</td>
<td>4.0000000</td>
<td>1.0000000</td>
<td>9.0000000</td>
<td>0</td>
<td>94.0000000</td>
</tr>
<tr>
<td>totCH14</td>
<td>32880</td>
<td>1.9705596</td>
<td>2.4737652</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>41.0000000</td>
</tr>
<tr>
<td>totCH15</td>
<td>32880</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>totCH16</td>
<td>32880</td>
<td>0.000942822</td>
<td>0.0306914</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
</tr>
<tr>
<td>totCH17</td>
<td>32880</td>
<td>0.0506995</td>
<td>0.2502156</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.0000000</td>
</tr>
<tr>
<td>totCH18</td>
<td>32880</td>
<td>7.2151460</td>
<td>6.2945968</td>
<td>6.0000000</td>
<td>3.0000000</td>
<td>10.0000000</td>
<td>0</td>
<td>64.0000000</td>
</tr>
<tr>
<td>totCH19</td>
<td>32880</td>
<td>2.2031022</td>
<td>3.6754148</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>56.0000000</td>
</tr>
<tr>
<td>totCH20</td>
<td>32880</td>
<td>0.0425182</td>
<td>0.2147684</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.0000000</td>
</tr>
<tr>
<td>totCH21</td>
<td>32880</td>
<td>4.8058090</td>
<td>3.2058311</td>
<td>4.0000000</td>
<td>2.0000000</td>
<td>7.0000000</td>
<td>0</td>
<td>29.0000000</td>
</tr>
<tr>
<td>totRad</td>
<td>32880</td>
<td>7.9857968</td>
<td>7.6682292</td>
<td>6.0000000</td>
<td>2.0000000</td>
<td>11.0000000</td>
<td>0</td>
<td>92.0000000</td>
</tr>
<tr>
<td>totAnesth</td>
<td>32880</td>
<td>0.0074209</td>
<td>0.1100488</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.0000000</td>
</tr>
<tr>
<td>TotEvalMan</td>
<td>32880</td>
<td>23.5829684</td>
<td>16.3484254</td>
<td>20.0000000</td>
<td>12.0000000</td>
<td>32.0000000</td>
<td>0</td>
<td>176.0000000</td>
</tr>
<tr>
<td>TotPathLab</td>
<td>32880</td>
<td>7.6703163</td>
<td>6.2565869</td>
<td>6.0000000</td>
<td>3.0000000</td>
<td>11.0000000</td>
<td>0</td>
<td>58.0000000</td>
</tr>
<tr>
<td>TotMed</td>
<td>32880</td>
<td>10.6562652</td>
<td>7.5360123</td>
<td>9.0000000</td>
<td>5.0000000</td>
<td>15.0000000</td>
<td>0</td>
<td>72.0000000</td>
</tr>
</tbody>
</table>
## Summary Statistics For Patients Who Escalate to MRI, Surgery, or Multiple Escalations

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Median</th>
<th>Lower Quartile</th>
<th>Upper Quartile</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>num_diag</td>
<td>13573</td>
<td>58.86888573</td>
<td>30.8133153</td>
<td>54.0000000</td>
<td>37.0000000</td>
<td>75.0000000</td>
<td>3.0000000</td>
<td>299.0000000</td>
</tr>
<tr>
<td>totcpt</td>
<td>13573</td>
<td>135.8301776</td>
<td>77.7860358</td>
<td>121.0000000</td>
<td>80.0000000</td>
<td>173.0000000</td>
<td>4.0000000</td>
<td>740.0000000</td>
</tr>
<tr>
<td>wscore</td>
<td>13573</td>
<td>3.3630001</td>
<td>2.8205555</td>
<td>3.0000000</td>
<td>1.0000000</td>
<td>5.0000000</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totcost</td>
<td>13573</td>
<td>617.3363</td>
<td>908.9701</td>
<td>332.4646</td>
<td>1562.67</td>
<td>7314.24</td>
<td>3932.10</td>
<td>2920390.37</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>13573</td>
<td>9.1894938</td>
<td>6.3343374</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>13.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH1</td>
<td>13573</td>
<td>1.2981655</td>
<td>1.5494198</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>14.0000000</td>
</tr>
<tr>
<td>totCH2</td>
<td>13573</td>
<td>1.8251676</td>
<td>2.5630298</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>25.0000000</td>
</tr>
<tr>
<td>totCH3</td>
<td>13573</td>
<td>0.6926251</td>
<td>1.2497460</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>13.0000000</td>
</tr>
<tr>
<td>totCH4</td>
<td>13573</td>
<td>4.1276063</td>
<td>3.6466072</td>
<td>3.0000000</td>
<td>2.0000000</td>
<td>6.0000000</td>
<td>0</td>
<td>42.0000000</td>
</tr>
<tr>
<td>totCH5</td>
<td>13573</td>
<td>0.8720990</td>
<td>1.7298366</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>22.0000000</td>
</tr>
<tr>
<td>totCH6</td>
<td>13573</td>
<td>1.7158329</td>
<td>2.0260783</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totCH7</td>
<td>13573</td>
<td>2.8435129</td>
<td>3.1628324</td>
<td>2.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>28.0000000</td>
</tr>
<tr>
<td>totCH8</td>
<td>13573</td>
<td>0.7595963</td>
<td>1.4740251</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totCH9</td>
<td>13573</td>
<td>5.9038532</td>
<td>5.3834373</td>
<td>4.0000000</td>
<td>2.0000000</td>
<td>8.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH10</td>
<td>13573</td>
<td>2.8420393</td>
<td>3.5274317</td>
<td>2.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>53.0000000</td>
</tr>
<tr>
<td>totCH11</td>
<td>13573</td>
<td>2.2045237</td>
<td>3.0161204</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>46.0000000</td>
</tr>
<tr>
<td>totCH12</td>
<td>13573</td>
<td>2.3205629</td>
<td>2.8212050</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>39.0000000</td>
</tr>
<tr>
<td>totCH13</td>
<td>13573</td>
<td>11.2957342</td>
<td>7.7794672</td>
<td>9.0000000</td>
<td>6.0000000</td>
<td>15.0000000</td>
<td>1.0000000</td>
<td>94.0000000</td>
</tr>
<tr>
<td>totCH14</td>
<td>13573</td>
<td>2.3066382</td>
<td>2.6352282</td>
<td>2.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>31.0000000</td>
</tr>
<tr>
<td>totCH15</td>
<td>13573</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>totCH16</td>
<td>13573</td>
<td>0.00073675</td>
<td>0.0271343</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
</tr>
<tr>
<td>totCH17</td>
<td>13573</td>
<td>0.0629927</td>
<td>0.2793541</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.0000000</td>
</tr>
<tr>
<td>totCH18</td>
<td>13573</td>
<td>8.7646062</td>
<td>6.8426104</td>
<td>7.0000000</td>
<td>4.0000000</td>
<td>12.0000000</td>
<td>0</td>
<td>64.0000000</td>
</tr>
<tr>
<td>totCH19</td>
<td>13573</td>
<td>2.9951374</td>
<td>4.2800203</td>
<td>1.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>56.0000000</td>
</tr>
<tr>
<td>totCH20</td>
<td>13573</td>
<td>0.0512046</td>
<td>0.2346704</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.0000000</td>
</tr>
<tr>
<td>totCH21</td>
<td>13573</td>
<td>5.9812127</td>
<td>3.3740463</td>
<td>6.0000000</td>
<td>4.0000000</td>
<td>8.0000000</td>
<td>0</td>
<td>29.0000000</td>
</tr>
<tr>
<td>totRad</td>
<td>13573</td>
<td>10.8947911</td>
<td>8.4846204</td>
<td>9.0000000</td>
<td>5.0000000</td>
<td>14.0000000</td>
<td>0</td>
<td>92.0000000</td>
</tr>
<tr>
<td>totAnesth</td>
<td>13573</td>
<td>0.0150298</td>
<td>0.1584941</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.0000000</td>
</tr>
<tr>
<td>TotEvalMan</td>
<td>13573</td>
<td>30.5824799</td>
<td>17.4825865</td>
<td>27.0000000</td>
<td>18.0000000</td>
<td>39.0000000</td>
<td>1.0000000</td>
<td>176.0000000</td>
</tr>
<tr>
<td>TotPathLab</td>
<td>13573</td>
<td>9.2380461</td>
<td>6.7204389</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>13.0000000</td>
<td>0</td>
<td>58.0000000</td>
</tr>
<tr>
<td>TotMed</td>
<td>13573</td>
<td>13.2264054</td>
<td>7.9124076</td>
<td>12.0000000</td>
<td>8.0000000</td>
<td>17.0000000</td>
<td>0</td>
<td>66.0000000</td>
</tr>
<tr>
<td>Variable</td>
<td>N</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Median</td>
<td>Lower Quartile</td>
<td>Upper Quartile</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>----------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>num_diag</td>
<td>19307</td>
<td>36.3064174</td>
<td>24.10939494</td>
<td>32.0000000</td>
<td>19.0000000</td>
<td>49.0000000</td>
<td>1.0000000</td>
<td>212.0000000</td>
</tr>
<tr>
<td>totcpt</td>
<td>19307</td>
<td>84.0360132</td>
<td>61.19520908</td>
<td>70.0000000</td>
<td>40.0000000</td>
<td>113.0000000</td>
<td>0</td>
<td>575.0000000</td>
</tr>
<tr>
<td>wscore</td>
<td>19307</td>
<td>3.0396814</td>
<td>2.7477762</td>
<td>2.0000000</td>
<td>1.0000000</td>
<td>4.0000000</td>
<td>0</td>
<td>21.0000000</td>
</tr>
<tr>
<td>totcost</td>
<td>19307</td>
<td>41086.23</td>
<td>71123.11</td>
<td>15422.95</td>
<td>5888.63</td>
<td>46317.09</td>
<td>23.170000</td>
<td>2228354.49</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>19307</td>
<td>6.1386544</td>
<td>5.1988722</td>
<td>5.0000000</td>
<td>2.0000000</td>
<td>9.0000000</td>
<td>0</td>
<td>42.0000000</td>
</tr>
<tr>
<td>totCH1</td>
<td>19307</td>
<td>0.8891594</td>
<td>1.2800779</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>12.0000000</td>
</tr>
<tr>
<td>totCH2</td>
<td>19307</td>
<td>1.3798222</td>
<td>2.3138910</td>
<td>0</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>30.0000000</td>
</tr>
<tr>
<td>totCH3</td>
<td>19307</td>
<td>0.5196997</td>
<td>1.0699440</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>13.0000000</td>
</tr>
<tr>
<td>totCH4</td>
<td>19307</td>
<td>3.1807635</td>
<td>3.1351770</td>
<td>2.0000000</td>
<td>1.0000000</td>
<td>4.0000000</td>
<td>0</td>
<td>34.0000000</td>
</tr>
<tr>
<td>totCH5</td>
<td>19307</td>
<td>0.6315844</td>
<td>1.5140226</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>51.0000000</td>
</tr>
<tr>
<td>totCH6</td>
<td>19307</td>
<td>0.8816491</td>
<td>1.4444471</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>18.0000000</td>
</tr>
<tr>
<td>totCH7</td>
<td>19307</td>
<td>2.0630341</td>
<td>2.7767915</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>28.0000000</td>
</tr>
<tr>
<td>totCH8</td>
<td>19307</td>
<td>0.4634588</td>
<td>1.1441474</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19.0000000</td>
</tr>
<tr>
<td>totCH9</td>
<td>19307</td>
<td>5.0855130</td>
<td>4.9920781</td>
<td>3.0000000</td>
<td>2.0000000</td>
<td>7.0000000</td>
<td>0</td>
<td>50.0000000</td>
</tr>
<tr>
<td>totCH10</td>
<td>19307</td>
<td>2.0747397</td>
<td>3.0116074</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>35.0000000</td>
</tr>
<tr>
<td>totCH11</td>
<td>19307</td>
<td>1.4067437</td>
<td>2.3416475</td>
<td>0</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>28.0000000</td>
</tr>
<tr>
<td>totCH12</td>
<td>19307</td>
<td>1.3884083</td>
<td>2.0436337</td>
<td>1.0000000</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>24.0000000</td>
</tr>
<tr>
<td>totCH13</td>
<td>19307</td>
<td>2.7738644</td>
<td>3.2536243</td>
<td>2.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>33.0000000</td>
</tr>
<tr>
<td>totCH14</td>
<td>19307</td>
<td>1.7342933</td>
<td>2.3247987</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>41.0000000</td>
</tr>
<tr>
<td>totCH15</td>
<td>19307</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>totCH16</td>
<td>19307</td>
<td>0.0010877</td>
<td>0.0329630</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
</tr>
<tr>
<td>totCH17</td>
<td>19307</td>
<td>0.0420573</td>
<td>0.2271176</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.0000000</td>
</tr>
<tr>
<td>totCH18</td>
<td>19307</td>
<td>6.1258611</td>
<td>5.6292446</td>
<td>5.0000000</td>
<td>2.0000000</td>
<td>9.0000000</td>
<td>0</td>
<td>48.0000000</td>
</tr>
<tr>
<td>totCH19</td>
<td>19307</td>
<td>1.6462941</td>
<td>3.0621818</td>
<td>0</td>
<td>0</td>
<td>2.0000000</td>
<td>0</td>
<td>39.0000000</td>
</tr>
<tr>
<td>totCH20</td>
<td>19307</td>
<td>0.0364117</td>
<td>0.1993726</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.0000000</td>
</tr>
<tr>
<td>totCH21</td>
<td>19307</td>
<td>3.9794893</td>
<td>2.8010065</td>
<td>4.0000000</td>
<td>2.0000000</td>
<td>6.0000000</td>
<td>0</td>
<td>22.0000000</td>
</tr>
<tr>
<td>totRad</td>
<td>19307</td>
<td>5.9407469</td>
<td>6.2771475</td>
<td>4.0000000</td>
<td>1.0000000</td>
<td>8.0000000</td>
<td>0</td>
<td>63.0000000</td>
</tr>
<tr>
<td>totAnesth</td>
<td>19307</td>
<td>0.0020718</td>
<td>0.0538178</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.0000000</td>
</tr>
<tr>
<td>TotEvalMan</td>
<td>19307</td>
<td>18.6622469</td>
<td>13.4778317</td>
<td>16.0000000</td>
<td>9.0000000</td>
<td>25.0000000</td>
<td>0</td>
<td>123.0000000</td>
</tr>
<tr>
<td>TotPathLab</td>
<td>19307</td>
<td>6.5681877</td>
<td>5.6581680</td>
<td>5.0000000</td>
<td>2.0000000</td>
<td>9.0000000</td>
<td>0</td>
<td>48.0000000</td>
</tr>
<tr>
<td>TotMed</td>
<td>19307</td>
<td>8.8494328</td>
<td>6.9311956</td>
<td>7.0000000</td>
<td>4.0000000</td>
<td>12.0000000</td>
<td>0</td>
<td>72.0000000</td>
</tr>
<tr>
<td>Variable</td>
<td>N</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Median</td>
<td>Lower Quartile</td>
<td>Upper Quartile</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>num_diag</td>
<td>2675</td>
<td>26.7086358</td>
<td>50.0000000</td>
<td>35.0000000</td>
<td>70.0000000</td>
<td>5.0000000</td>
<td>228.0000000</td>
<td></td>
</tr>
<tr>
<td>totcpt</td>
<td>2675</td>
<td>74.3438340</td>
<td>112.0000000</td>
<td>74.0000000</td>
<td>161.0000000</td>
<td>8.0000000</td>
<td>689.0000000</td>
<td></td>
</tr>
<tr>
<td>wscore</td>
<td>2675</td>
<td>2.9908535</td>
<td>3.0000000</td>
<td>1.0000000</td>
<td>5.0000000</td>
<td>0</td>
<td>17.0000000</td>
<td></td>
</tr>
<tr>
<td>totcost</td>
<td>2675</td>
<td>28058.75</td>
<td>28058.75</td>
<td>12795.49</td>
<td>67678.16</td>
<td>672.1500000</td>
<td>1661196.84</td>
<td></td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>2675</td>
<td>5.9686968</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>12.0000000</td>
<td>0</td>
<td>45.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH1</td>
<td>2675</td>
<td>1.5578844</td>
<td>1.0000000</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>17.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH2</td>
<td>2675</td>
<td>2.5924817</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>24.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH3</td>
<td>2675</td>
<td>1.2196791</td>
<td>1.0000000</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>11.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH4</td>
<td>2675</td>
<td>3.7996708</td>
<td>3.0000000</td>
<td>1.0000000</td>
<td>6.0000000</td>
<td>0</td>
<td>42.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH5</td>
<td>2675</td>
<td>1.7994083</td>
<td>1.0000000</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>17.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH6</td>
<td>2675</td>
<td>1.9949532</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>19.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH7</td>
<td>2675</td>
<td>2.9737173</td>
<td>2.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>27.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH8</td>
<td>2675</td>
<td>1.3873467</td>
<td>1.0000000</td>
<td>0</td>
<td>1.0000000</td>
<td>0</td>
<td>11.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH9</td>
<td>2675</td>
<td>4.0000000</td>
<td>4.0000000</td>
<td>2.0000000</td>
<td>8.0000000</td>
<td>0</td>
<td>44.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH10</td>
<td>2675</td>
<td>3.6009086</td>
<td>2.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>34.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH11</td>
<td>2675</td>
<td>2.9784012</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>27.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH12</td>
<td>2675</td>
<td>2.6107207</td>
<td>1.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>28.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH13</td>
<td>2675</td>
<td>5.4284144</td>
<td>8.0000000</td>
<td>5.0000000</td>
<td>12.0000000</td>
<td>0</td>
<td>43.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH14</td>
<td>2675</td>
<td>2.7129301</td>
<td>2.0000000</td>
<td>0</td>
<td>3.0000000</td>
<td>0</td>
<td>31.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH15</td>
<td>2675</td>
<td>0.0000000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>totCH16</td>
<td>2675</td>
<td>0.0386478</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH17</td>
<td>2675</td>
<td>0.2791180</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH18</td>
<td>2675</td>
<td>6.9780414</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>13.0000000</td>
<td>0</td>
<td>52.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH19</td>
<td>2675</td>
<td>4.028131</td>
<td>1.0000000</td>
<td>0</td>
<td>4.0000000</td>
<td>0</td>
<td>37.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH20</td>
<td>2675</td>
<td>0.2161837</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.0000000</td>
<td></td>
</tr>
<tr>
<td>totCH21</td>
<td>2675</td>
<td>3.1057992</td>
<td>5.0000000</td>
<td>3.0000000</td>
<td>7.0000000</td>
<td>0</td>
<td>24.0000000</td>
<td></td>
</tr>
<tr>
<td>totRad</td>
<td>2675</td>
<td>8.1118522</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>13.0000000</td>
<td>0</td>
<td>84.0000000</td>
<td></td>
</tr>
<tr>
<td>totAnesth</td>
<td>2675</td>
<td>0.1607125</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.0000000</td>
<td></td>
</tr>
<tr>
<td>TotPathLab</td>
<td>2675</td>
<td>8.4548348</td>
<td>8.0000000</td>
<td>4.0000000</td>
<td>13.0000000</td>
<td>0</td>
<td>42.0000000</td>
<td></td>
</tr>
<tr>
<td>TotMed</td>
<td>2675</td>
<td>7.3686271</td>
<td>11.0000000</td>
<td>7.0000000</td>
<td>16.0000000</td>
<td>0</td>
<td>66.0000000</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>N</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Median</td>
<td>Lower Quartile</td>
<td>Upper Quartile</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>num_diag</td>
<td>6606</td>
<td>52.9279</td>
<td>122.22</td>
<td>3.21</td>
<td>55.3470</td>
<td>8.44</td>
<td>0</td>
<td>4.00</td>
</tr>
<tr>
<td>totcpt</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>wscore</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totcost</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH1</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH2</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH3</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH4</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH5</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH6</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH7</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH8</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH9</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH10</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH11</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH12</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH13</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH14</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH15</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH16</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH17</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH18</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH19</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH20</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totCH21</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totRad</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>totAnesth</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>TotEvalMan</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>TotPathLab</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
<tr>
<td>TotMed</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
<td>6606</td>
</tr>
</tbody>
</table>
APPENDIX 4

Machine Learning Modeling

Class Bias

Before running any machine learning models, AERG checked for class bias within our data set. Class bias occurs when the class distributions within a data set are highly imbalanced. More specifically, class bias occurs when the proportion of events is much smaller than the proportion of non-events.

For example, in our data the proportion of patients that escalated was far less than the proportion of patients that did not escalate; 13,573 patients had an “Escalation” while 19,307 patients had “No Escalation.” Similarly, the proportion of patients that had a “Surgery without MRI” was far less than the proportion of patients that had “No Escalation.” The largest class bias was among the proportion of patients with a “MRI without surgery” versus the proportion of patients with “No Escalation.” There are several ways to account for this imbalance such as Up-Sampling, Down Sampling, collecting more data, resampling, synthetic sampling (SMOTE). We choose to address the issue of class bias when making our training and testing data sets.

Training and Test Sets

Typically, for predictive modeling, the data is split into a training set and a testing set. Most commonly the data is split in which 80% of observations are assigned to the training set and 20% of observations are assigned to the testing set.
Hyper parameter Tuning

Machine learning requires tuning of the parameters in order to perform optimally. There are different parameters with different algorithms. For example, tree-based models have a number of branches and neural networks have a number of hidden layers. These are set initially by the researcher for the initial test. The learning occurs when the parameters are tuned. This is a procedure which occurs through trial and error. What occurs internally is the parameters are changed, or tuned, and the data is run through again. Once the new run completes it compares the performance on validation test sets to determine if the tunes are better or worse than previous runs. This tuning process finds the optimal model within the machine learning nomenclature.

In Azure we use the module called Tune Model Hyper parameter. It is an all-encompassing step in the Azure platform to assist the researcher with tuning the model for optimal performance. The inputs for this step are how many sweeps of the data the researcher would prefer and what the sweeps are tuning for. Recall that the accuracy metrics are used to recognize a number of different outcome variables. For example, if the research was to identify overall accuracy – predicting both positive and negative outcomes – we would use the accuracy metric for tuning. This is what we did in our models.

Overfitting

Overfitting is a concern when using machine learning. Since the models are so intricate and track several details among the data, each iteration contains noise as well as signal. Noise within a dataset affects the predictions. Predicting generally as well as predicting specific observations is not as easy to do because of the nature of the data. For example, if a particular dataset contains much more of one particular outcome as opposed to another it can cause issues
with the results. This is referred to as a class imbalance or class bias. More on this later in this report.

Essentially, if a model is over fit, the function within the algorithm too closely fits the data which serves as the model's inputs. When new data applied to the model, overfitting causes the model to not predict as well. The way that you correct for overfitting within machine learning is to split the data into two different sets. One data set is used for training the model and the other data set is used for testing the model. This allows for the model accuracy to be truly accurate. In both our Azure and R machine learning algorithms we took this approach to overfitting.

**Model Validation**

Our models used two different outcome variables so we evaluated the models a little differently. For the models which use a binary model we use the evaluate module which is pretty straightforward. It reviews the model with new, unseen data. It provides a confusion matrix with some additional measurement metrics. This output is an easy way quickly evaluate the model.

<table>
<thead>
<tr>
<th>True Positive</th>
<th>False Negative</th>
<th>Accuracy</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1698</td>
<td>1042</td>
<td>0.784</td>
<td>0.819</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>False Positive</th>
<th>True Negative</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>376</td>
<td>3460</td>
<td>0.620</td>
<td>0.705</td>
</tr>
</tbody>
</table>

Cross validation is a technique used within the training and testing portion of machine learning. It is both an evaluation and tuning feature of machine learning. After a model is trained it needs to test the parameters on new data. This cross validation technique uses subsets of the data called folds and runs the models on these folds and then validates on a validation fold. In
Azure we randomly broke our data into 10 folds of our dataset. One fold was held out as a validation fold, it is known as the holdout fold. The cross validation module evaluates the model on similar metrics as the evaluation module. Precision, recall, F-score, AUC, average log loss, training log loss are all measured with this module and Azure picks the best model among these as its highest preforming model.

**Variable Importance Factor**

It is important to look at the relative importance of each variable in terms of its contribution to the escalation categories. This analysis does not make any assumptions about the practical significance of a variable but simply attempts to explain a variable’s contribution to the predictable variation of a model. This examination also adds additional support when checking for multicollinearity amongst variables.

**Misclassification Error**

The misclassification issue is a frequent hurdle for most models when dealing with categorical dependent variables which can lead to a biased estimate. A misclassification error occurs when we know a specific individual should be placed into a certain class but is not by the model. The overall goal is to minimize the misclassification error as much as possible. However, it is important to note that not all misclassifications are the same, you might find certain variable combinations are more important (predictive importance) than others.
Receiver Operating Characteristics Curve (ROC)

The ROC curve displays how accurately our logit model is predicting true outcomes. Meaning, it is measuring how well we are predicting true 1s in our model compared to false positives at various designated thresholds.

The ROC curve plots the Sensitivity (true positive rate) and the Specificity (false positive rate). Each point on the ROC curve represents a pair of Sensitivity and Specificity that corresponds to a particular decision threshold. The ROC measures how the model is able to predict a categorical dependent variable.
APPENDIX 5

Two-Class Boosted Decision Tree Results: Predicting Escalation

For the Two-Class Boosted Decision Tree model predicting escalation, the overall accuracy is approximately 82%. The percent of correctly predicted escalations (sensitivity) is approximately 78%. The percent of correctly predicted non-escalations is approximately 78%.

Figure: ROC Curve
Two-Class Boosted Decision Tree Results: Predicting Escalation

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>Weight</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTCH13</td>
<td>0.335341</td>
<td>1</td>
</tr>
<tr>
<td>TOTCPT</td>
<td>0.113564</td>
<td>2</td>
</tr>
<tr>
<td>NUM_DIAG</td>
<td>0.033577</td>
<td>3</td>
</tr>
<tr>
<td>TOTCH18</td>
<td>0.027068</td>
<td>4</td>
</tr>
<tr>
<td>TOTCOST</td>
<td>0.026764</td>
<td>5</td>
</tr>
<tr>
<td>TOTPATH_LAB</td>
<td>0.018187</td>
<td>6</td>
</tr>
<tr>
<td>SUMEVAL_MANAGE</td>
<td>0.014538</td>
<td>7</td>
</tr>
<tr>
<td>WSCORE</td>
<td>0.014477</td>
<td>8</td>
</tr>
<tr>
<td>TOTMEDICINE</td>
<td>0.011618</td>
<td>9</td>
</tr>
<tr>
<td>TOTCH19</td>
<td>0.011496</td>
<td>10</td>
</tr>
<tr>
<td>TOTCH9</td>
<td>0.011131</td>
<td>11</td>
</tr>
<tr>
<td>TOTRAD</td>
<td>0.009793</td>
<td>12</td>
</tr>
<tr>
<td>NON_MSK_SURG</td>
<td>0.008151</td>
<td>13</td>
</tr>
<tr>
<td>TOTCH21</td>
<td>0.007786</td>
<td>14</td>
</tr>
<tr>
<td>TOTCH4</td>
<td>0.007056</td>
<td>15</td>
</tr>
<tr>
<td>START_2015</td>
<td>0.006326</td>
<td>16</td>
</tr>
<tr>
<td>TOTCH14</td>
<td>0.006265</td>
<td>17</td>
</tr>
<tr>
<td>TOTCH5</td>
<td>0.006204</td>
<td>18</td>
</tr>
<tr>
<td>TOTCH2</td>
<td>0.006144</td>
<td>19</td>
</tr>
<tr>
<td>TOTCH11</td>
<td>0.006022</td>
<td>20</td>
</tr>
<tr>
<td>TOTCH6</td>
<td>0.005474</td>
<td>21</td>
</tr>
<tr>
<td>TOTCH7</td>
<td>0.005353</td>
<td>22</td>
</tr>
<tr>
<td>TOTCH10</td>
<td>0.005231</td>
<td>23</td>
</tr>
<tr>
<td>TOTCH12</td>
<td>0.004927</td>
<td>24</td>
</tr>
<tr>
<td>TOTCH1</td>
<td>0.004136</td>
<td>25</td>
</tr>
<tr>
<td>TOTCH3</td>
<td>0.003832</td>
<td>26</td>
</tr>
<tr>
<td>TOTCH8</td>
<td>0.003163</td>
<td>27</td>
</tr>
<tr>
<td>TOTDUR</td>
<td>0.002494</td>
<td>28</td>
</tr>
<tr>
<td>ONE_YEAR</td>
<td>0.000912</td>
<td>29</td>
</tr>
<tr>
<td>TOTCH17</td>
<td>0.000487</td>
<td>30</td>
</tr>
<tr>
<td>TOTCH20</td>
<td>0.000182</td>
<td>31</td>
</tr>
<tr>
<td>LESS_ONE_YEAR</td>
<td>0.000182</td>
<td>32</td>
</tr>
<tr>
<td>FOUR_YEARS</td>
<td>0.000182</td>
<td>33</td>
</tr>
</tbody>
</table>
Two-Class Boosted Decision Tree Results ~ Predicting a Surgery without MRI

For the Two-Class Boosted Decision Tree model predicting Surgery without MRI, the overall accuracy is approximately 80%. The percent of correctly predicted Surgeries without MRI (sensitivity) is approximately 66%. The percent of correctly predicted non-escalations is approximately 71%.

Figure: ROC Curve
Two-Class Boosted Decision Tree Results ~ Predicting a Surgery without MRI

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>WEIGHT</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>totCH13</td>
<td>0.214112</td>
<td>1</td>
</tr>
<tr>
<td>totCH21</td>
<td>0.005474</td>
<td>2</td>
</tr>
<tr>
<td>totCH18</td>
<td>0.005322</td>
<td>3</td>
</tr>
<tr>
<td>totCPT</td>
<td>0.00517</td>
<td>4</td>
</tr>
<tr>
<td>totcost</td>
<td>0.004866</td>
<td>5</td>
</tr>
<tr>
<td>totRad</td>
<td>0.003345</td>
<td>6</td>
</tr>
<tr>
<td>totCH5</td>
<td>0.002281</td>
<td>7</td>
</tr>
<tr>
<td>num_diag</td>
<td>0.001369</td>
<td>8</td>
</tr>
<tr>
<td>totMedicine</td>
<td>0.000912</td>
<td>9</td>
</tr>
<tr>
<td>totdur</td>
<td>0.000608</td>
<td>10</td>
</tr>
<tr>
<td>totPath_Lab</td>
<td>0.000456</td>
<td>11</td>
</tr>
<tr>
<td>sumEval_Manage</td>
<td>0.000304</td>
<td>12</td>
</tr>
<tr>
<td>three_years</td>
<td>0.000304</td>
<td>13</td>
</tr>
<tr>
<td>start_2017</td>
<td>0.000304</td>
<td>14</td>
</tr>
<tr>
<td>totCH14</td>
<td>0.000152</td>
<td>15</td>
</tr>
<tr>
<td>four_years</td>
<td>0.000152</td>
<td>16</td>
</tr>
<tr>
<td>start_2016</td>
<td>0.000152</td>
<td>17</td>
</tr>
</tbody>
</table>
Two-Class Boosted Decision Tree Results ~ Predicting MRI without surgery

For the Two-Class Boosted Decision Tree model predicting MRI without Surgery, the overall accuracy is approximately 83%. The percent of correctly predicted MRIs without Surgery (sensitivity) is approximately 48%. The percent of correctly predicted non-escalations is approximately 59%.

Figure: ROC Curve
Two-Class Boosted Decision Tree Results ~ Predicting MRI without surgery

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>Weight</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUM_DIAG</td>
<td>0.235401</td>
<td>1</td>
</tr>
<tr>
<td>TOTCH9</td>
<td>0.034824</td>
<td>2</td>
</tr>
<tr>
<td>TOTCH18</td>
<td>0.021137</td>
<td>3</td>
</tr>
<tr>
<td>TOTCH19</td>
<td>0.012318</td>
<td>4</td>
</tr>
<tr>
<td>TOTCH7</td>
<td>0.008364</td>
<td>5</td>
</tr>
<tr>
<td>TOTCH14</td>
<td>0.006691</td>
<td>6</td>
</tr>
<tr>
<td>TOTCH10</td>
<td>0.005779</td>
<td>7</td>
</tr>
<tr>
<td>TOTCH4</td>
<td>0.005474</td>
<td>8</td>
</tr>
<tr>
<td>TOTCPT</td>
<td>0.005322</td>
<td>9</td>
</tr>
<tr>
<td>TOTCH11</td>
<td>0.004866</td>
<td>10</td>
</tr>
<tr>
<td>TOTCH21</td>
<td>0.002737</td>
<td>11</td>
</tr>
<tr>
<td>TOTCH12</td>
<td>0.002585</td>
<td>12</td>
</tr>
<tr>
<td>TOTCH3</td>
<td>0.002281</td>
<td>13</td>
</tr>
<tr>
<td>TOTCH2</td>
<td>0.002129</td>
<td>14</td>
</tr>
<tr>
<td>TOTRAD</td>
<td>0.002129</td>
<td>15</td>
</tr>
<tr>
<td>TOTCH1</td>
<td>0.001825</td>
<td>16</td>
</tr>
<tr>
<td>TOTCOST</td>
<td>0.001521</td>
<td>17</td>
</tr>
<tr>
<td>TOTCH8</td>
<td>0.001217</td>
<td>18</td>
</tr>
<tr>
<td>TOTCH5</td>
<td>0.001064</td>
<td>19</td>
</tr>
<tr>
<td>TOTDUR</td>
<td>0.001064</td>
<td>20</td>
</tr>
<tr>
<td>WSCORE</td>
<td>0.000912</td>
<td>21</td>
</tr>
<tr>
<td>SUMEVAL_MANAGE</td>
<td>0.000912</td>
<td>22</td>
</tr>
<tr>
<td>TOTANESTH</td>
<td>0.000608</td>
<td>23</td>
</tr>
<tr>
<td>TOTPATH_LAB</td>
<td>0.000608</td>
<td>24</td>
</tr>
<tr>
<td>NON_MSK_SURG</td>
<td>0.000304</td>
<td>25</td>
</tr>
<tr>
<td>TOTCH20</td>
<td>0.000152</td>
<td>26</td>
</tr>
</tbody>
</table>
Two-Class Boosted Decision Tree Results ~ Predicting Multiple Escalation

For the Two-Class Boosted Decision Tree model predicting Multiple Escalation, the overall accuracy is approximately 89%. The percent of correctly predicted Multiple Escalations (sensitivity) is approximately 47%. The percent of correctly predicted non-escalations is approximately 57%.

Figure: ROC Curve
Two-Class Boosted Decision Tree Results ~ Predicting Multiple Escalation
<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>totCH13</td>
<td>0.08896</td>
<td>1</td>
</tr>
<tr>
<td>totCPT</td>
<td>0.005474</td>
<td>2</td>
</tr>
<tr>
<td>num_diag</td>
<td>0.00517</td>
<td>3</td>
</tr>
<tr>
<td>totcost</td>
<td>0.003498</td>
<td>4</td>
</tr>
<tr>
<td>sumEval_Manage</td>
<td>0.001521</td>
<td>5</td>
</tr>
<tr>
<td>three_years</td>
<td>0.001369</td>
<td>6</td>
</tr>
<tr>
<td>totCH7</td>
<td>0.001217</td>
<td>7</td>
</tr>
<tr>
<td>start_2015</td>
<td>0.001217</td>
<td>8</td>
</tr>
<tr>
<td>start_2017</td>
<td>0.001064</td>
<td>9</td>
</tr>
<tr>
<td>totCH1</td>
<td>0.00076</td>
<td>10</td>
</tr>
<tr>
<td>totCH12</td>
<td>0.00076</td>
<td>11</td>
</tr>
<tr>
<td>totCH14</td>
<td>0.000608</td>
<td>12</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>0.000456</td>
<td>13</td>
</tr>
<tr>
<td>totCH17</td>
<td>0.000456</td>
<td>14</td>
</tr>
<tr>
<td>totCH16</td>
<td>0.000152</td>
<td>15</td>
</tr>
<tr>
<td>totCH20</td>
<td>0.000152</td>
<td>16</td>
</tr>
<tr>
<td>totAnesth</td>
<td>0.000152</td>
<td>17</td>
</tr>
<tr>
<td>start_2016</td>
<td>0.000152</td>
<td>18</td>
</tr>
<tr>
<td>totCH5</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>totCH15</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>totCH19</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>totPath_Lab</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>totMedicine</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>less_one_year</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>start_2014</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>start_2018</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>totCH3</td>
<td>-0.00015</td>
<td>27</td>
</tr>
<tr>
<td>totCH8</td>
<td>-0.00015</td>
<td>28</td>
</tr>
<tr>
<td>totRad</td>
<td>-0.00015</td>
<td>29</td>
</tr>
<tr>
<td>two_years</td>
<td>-0.00015</td>
<td>30</td>
</tr>
<tr>
<td>totCH4</td>
<td>-0.0003</td>
<td>31</td>
</tr>
<tr>
<td>totCH21</td>
<td>-0.0003</td>
<td>32</td>
</tr>
<tr>
<td>one_year</td>
<td>-0.0003</td>
<td>33</td>
</tr>
<tr>
<td>four_years</td>
<td>-0.0003</td>
<td>34</td>
</tr>
<tr>
<td>totCH9</td>
<td>-0.00046</td>
<td>35</td>
</tr>
<tr>
<td>totCH6</td>
<td>-0.00061</td>
<td>36</td>
</tr>
<tr>
<td>Variable</td>
<td>Score</td>
<td>Line</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>totCH18</td>
<td>-0.00061</td>
<td>37</td>
</tr>
<tr>
<td>wscore</td>
<td>-0.00076</td>
<td>38</td>
</tr>
<tr>
<td>totCH11</td>
<td>-0.00106</td>
<td>39</td>
</tr>
<tr>
<td>totalur</td>
<td>-0.00106</td>
<td>40</td>
</tr>
<tr>
<td>totCH10</td>
<td>-0.00137</td>
<td>41</td>
</tr>
<tr>
<td>totCH2</td>
<td>-0.00152</td>
<td>42</td>
</tr>
</tbody>
</table>
# Multi-Class Random Forest Results ~ Predicting Escalation Category

## Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall accuracy</td>
<td>0.649027</td>
<td>0.678285</td>
</tr>
<tr>
<td>Average accuracy</td>
<td>0.824513</td>
<td>0.839142</td>
</tr>
<tr>
<td>Micro-averaged precision</td>
<td>0.649027</td>
<td>0.678285</td>
</tr>
<tr>
<td>Macro-averaged precision</td>
<td>0.445537</td>
<td>0.475569</td>
</tr>
<tr>
<td>Micro-averaged recall</td>
<td>0.649027</td>
<td>0.678285</td>
</tr>
<tr>
<td>Macro-averaged recall</td>
<td>0.445514</td>
<td>0.460494</td>
</tr>
</tbody>
</table>

## Confusion Matrix

**Predicted Class**

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>MRI with..</th>
<th>Multiple..</th>
<th>No Esca..</th>
<th>Surgery..</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI with..</td>
<td>9.3%</td>
<td>18.0%</td>
<td>47.6%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Multiple..</td>
<td>6.9%</td>
<td>55.0%</td>
<td>12.6%</td>
<td>25.6%</td>
</tr>
<tr>
<td>No Esca..</td>
<td>2.8%</td>
<td>2.5%</td>
<td>88.3%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Surgery..</td>
<td>8.3%</td>
<td>21.1%</td>
<td>45.0%</td>
<td>25.7%</td>
</tr>
</tbody>
</table>

**Actual Class**

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>MRI with..</th>
<th>Multiple..</th>
<th>No Esca..</th>
<th>Surgery..</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI with..</td>
<td>1.9%</td>
<td>13.2%</td>
<td>46.3%</td>
<td>33.5%</td>
</tr>
<tr>
<td>Multiple..</td>
<td>1.6%</td>
<td>58.2%</td>
<td>11.7%</td>
<td>30.5%</td>
</tr>
<tr>
<td>No Esca..</td>
<td>0.4%</td>
<td>1.7%</td>
<td>90.6%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Surgery..</td>
<td>1.4%</td>
<td>17.8%</td>
<td>45.3%</td>
<td>35.5%</td>
</tr>
</tbody>
</table>
Two-Class Neural Network Results ~ Predicting Escalation

For the Two-Class Neural Network model predicting escalation, the overall accuracy is approximately 82%. The percent of correctly predicted escalations (sensitivity) is approximately 72%. The percent of correctly predicted non-escalations is approximately 72%.

Figure: ROC Curve
Two-Class Neural Network Results ~ Predicting Escalation

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>Weight</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTCH13</td>
<td>0.268613</td>
<td>1</td>
</tr>
<tr>
<td>TOTCPT</td>
<td>0.052311</td>
<td>2</td>
</tr>
<tr>
<td>NUM_DIAG</td>
<td>0.028285</td>
<td>3</td>
</tr>
<tr>
<td>TOTCH18</td>
<td>0.015633</td>
<td>4</td>
</tr>
<tr>
<td>TOTCH9</td>
<td>0.006813</td>
<td>5</td>
</tr>
<tr>
<td>TOTCH19</td>
<td>0.006813</td>
<td>6</td>
</tr>
<tr>
<td>TOTPATH_LAB</td>
<td>0.005596</td>
<td>7</td>
</tr>
<tr>
<td>WSCORE</td>
<td>0.003589</td>
<td>8</td>
</tr>
<tr>
<td>TOTCH10</td>
<td>0.003528</td>
<td>9</td>
</tr>
<tr>
<td>SUMEVAL_MANAGE</td>
<td>0.003467</td>
<td>10</td>
</tr>
<tr>
<td>TOTCH5</td>
<td>0.00292</td>
<td>11</td>
</tr>
<tr>
<td>TOTCH7</td>
<td>0.00292</td>
<td>12</td>
</tr>
<tr>
<td>ONE_YEAR</td>
<td>0.002798</td>
<td>13</td>
</tr>
<tr>
<td>NON_MSK_SURG</td>
<td>0.002494</td>
<td>14</td>
</tr>
<tr>
<td>TOTCH1</td>
<td>0.002494</td>
<td>15</td>
</tr>
<tr>
<td>TOTCH14</td>
<td>0.002433</td>
<td>16</td>
</tr>
<tr>
<td>TOTCH12</td>
<td>0.002372</td>
<td>17</td>
</tr>
<tr>
<td>TOTDUR</td>
<td>0.001399</td>
<td>18</td>
</tr>
<tr>
<td>TOTCH11</td>
<td>0.001277</td>
<td>19</td>
</tr>
<tr>
<td>START_2015</td>
<td>0.001277</td>
<td>20</td>
</tr>
<tr>
<td>TOTCH3</td>
<td>0.001217</td>
<td>21</td>
</tr>
<tr>
<td>START_2017</td>
<td>0.001217</td>
<td>22</td>
</tr>
<tr>
<td>THREE_YEARS</td>
<td>0.00073</td>
<td>23</td>
</tr>
<tr>
<td>TOTCH2</td>
<td>0.000669</td>
<td>24</td>
</tr>
<tr>
<td>TOTCH6</td>
<td>0.000365</td>
<td>25</td>
</tr>
<tr>
<td>TOTCH21</td>
<td>0.000365</td>
<td>26</td>
</tr>
<tr>
<td>TOTCH17</td>
<td>0.000304</td>
<td>27</td>
</tr>
<tr>
<td>LESS_ONE_YEAR</td>
<td>0.000304</td>
<td>28</td>
</tr>
<tr>
<td>TWO_YEARS</td>
<td>0.000304</td>
<td>29</td>
</tr>
<tr>
<td>TOTCH8</td>
<td>0.000243</td>
<td>30</td>
</tr>
<tr>
<td>START_2016</td>
<td>0.000243</td>
<td>31</td>
</tr>
<tr>
<td>TOTRAD</td>
<td>0.000122</td>
<td>32</td>
</tr>
<tr>
<td>START_2018</td>
<td>6.08E-05</td>
<td>33</td>
</tr>
</tbody>
</table>
Two-Class Neural Network Results ~ Predicting a Surgery without MRI

For the Two Class Neural Network model predicting Surgery without MRI, the overall accuracy is approximately 80%. The percent of correctly predicted Surgeries without MRI (sensitivity) is approximately 68%. The percent of correctly predicted non-escalations is approximately 71%.

Figure: ROC Curve
### VARIABLE IMPORTANCE

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>Weight</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTCH13</td>
<td>0.19191</td>
<td>1</td>
</tr>
<tr>
<td>TOTCPT</td>
<td>0.028437</td>
<td>2</td>
</tr>
<tr>
<td>SUMEVAL_MANAGE</td>
<td>0.021594</td>
<td>3</td>
</tr>
<tr>
<td>TOTCH18</td>
<td>0.020377</td>
<td>4</td>
</tr>
<tr>
<td>TOTCH19</td>
<td>0.005627</td>
<td>5</td>
</tr>
<tr>
<td>TOTRAD</td>
<td>0.004258</td>
<td>6</td>
</tr>
<tr>
<td>TOTCH9</td>
<td>0.002129</td>
<td>7</td>
</tr>
<tr>
<td>TOTCH5</td>
<td>0.001825</td>
<td>8</td>
</tr>
<tr>
<td>TOTCH10</td>
<td>0.001825</td>
<td>9</td>
</tr>
<tr>
<td>TOTPATH_LAB</td>
<td>0.001825</td>
<td>10</td>
</tr>
<tr>
<td>START_2015</td>
<td>0.001521</td>
<td>11</td>
</tr>
<tr>
<td>START_2016</td>
<td>0.000608</td>
<td>12</td>
</tr>
<tr>
<td>START_2017</td>
<td>0.000608</td>
<td>13</td>
</tr>
<tr>
<td>TOTCH7</td>
<td>0.000456</td>
<td>14</td>
</tr>
<tr>
<td>NUM_DIAG</td>
<td>0.000304</td>
<td>15</td>
</tr>
<tr>
<td>TOTCH8</td>
<td>0.000304</td>
<td>16</td>
</tr>
<tr>
<td>TOTCH16</td>
<td>0.000304</td>
<td>17</td>
</tr>
<tr>
<td>TOTCH12</td>
<td>0.000152</td>
<td>18</td>
</tr>
</tbody>
</table>
Two-Class Neural Network Results ~ Predicting a MRI without surgery

For the Two Class Neural Network model predicting MRI without Surgery, the overall accuracy is approximately 84%. The percent of correctly predicted Surgeries without MRI (sensitivity) is approximately 51%. The percent of correctly predicted non-escalations is approximately 61%.

Figure: ROC Curve
Two-Class Neural Network Results ~ Predicting a MRI without surgery

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>SCORE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>totCH13</td>
<td>0.145985</td>
<td>1</td>
</tr>
<tr>
<td>totdur</td>
<td>0.001977</td>
<td>2</td>
</tr>
<tr>
<td>wscore</td>
<td>0.001825</td>
<td>3</td>
</tr>
<tr>
<td>num_diag</td>
<td>0.001369</td>
<td>4</td>
</tr>
<tr>
<td>start_2015</td>
<td>0.001369</td>
<td>5</td>
</tr>
<tr>
<td>totcost</td>
<td>0.001217</td>
<td>6</td>
</tr>
<tr>
<td>one_year</td>
<td>0.001064</td>
<td>7</td>
</tr>
<tr>
<td>totCH21</td>
<td>0.000912</td>
<td>8</td>
</tr>
<tr>
<td>totCH3</td>
<td>0.000608</td>
<td>9</td>
</tr>
<tr>
<td>totMedicine</td>
<td>0.000456</td>
<td>10</td>
</tr>
<tr>
<td>start_2017</td>
<td>0.000456</td>
<td>11</td>
</tr>
<tr>
<td>totCH7</td>
<td>0.000304</td>
<td>12</td>
</tr>
<tr>
<td>totCH10</td>
<td>0.000304</td>
<td>13</td>
</tr>
<tr>
<td>totCH20</td>
<td>0.000304</td>
<td>14</td>
</tr>
<tr>
<td>totAnesth</td>
<td>0.000304</td>
<td>15</td>
</tr>
<tr>
<td>sumEval_Manage</td>
<td>0.000304</td>
<td>16</td>
</tr>
<tr>
<td>start_2018</td>
<td>0.000304</td>
<td>17</td>
</tr>
<tr>
<td>totCPT</td>
<td>0.000304</td>
<td>18</td>
</tr>
<tr>
<td>totCH2</td>
<td>0.000152</td>
<td>19</td>
</tr>
<tr>
<td>totCH12</td>
<td>0.000152</td>
<td>20</td>
</tr>
<tr>
<td>totCH17</td>
<td>0.000152</td>
<td>21</td>
</tr>
<tr>
<td>totCH18</td>
<td>0.000152</td>
<td>22</td>
</tr>
<tr>
<td>totCH19</td>
<td>0.000152</td>
<td>23</td>
</tr>
<tr>
<td>four_years</td>
<td>0.000152</td>
<td>24</td>
</tr>
<tr>
<td>totCH4</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>totCH14</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>totCH15</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>totCH16</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>less_one_year</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>two_years</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>three_years</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>start_2014</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>start_2016</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>totCH6</td>
<td>-0.00015</td>
<td>34</td>
</tr>
<tr>
<td>totCH8</td>
<td>-0.00015</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>totRad</td>
<td>-0.00015</td>
<td>36</td>
</tr>
<tr>
<td>totCH5</td>
<td>-0.00046</td>
<td>37</td>
</tr>
<tr>
<td>totCH1</td>
<td>-0.00061</td>
<td>38</td>
</tr>
<tr>
<td>non_msk_surg</td>
<td>-0.00076</td>
<td>39</td>
</tr>
<tr>
<td>totPath_Lab</td>
<td>-0.00091</td>
<td>40</td>
</tr>
<tr>
<td>totCH9</td>
<td>-0.00167</td>
<td>41</td>
</tr>
<tr>
<td>totCH11</td>
<td>-0.00182</td>
<td>42</td>
</tr>
</tbody>
</table>
Two-Class Neural Network Results ~ Predicting Multiple Escalation

For the Two Class Neural Network model predicting multiple escalation, the overall accuracy is approximately 89%. The percent of correctly predicted multiple escalations (sensitivity) is approximately 58%. The percent of correctly predicted non-escalations is approximately 55%.
## Two-Class Neural Network Results ~ Predicting Multiple Escalation

<table>
<thead>
<tr>
<th>VARIABLE IMPORTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEATURE</td>
</tr>
<tr>
<td>num_diag</td>
</tr>
<tr>
<td>totCH18</td>
</tr>
<tr>
<td>totCPT</td>
</tr>
<tr>
<td>totCH9</td>
</tr>
<tr>
<td>totCH4</td>
</tr>
<tr>
<td>totCH7</td>
</tr>
<tr>
<td>totCH19</td>
</tr>
<tr>
<td>totCH21</td>
</tr>
<tr>
<td>totCH14</td>
</tr>
<tr>
<td>totCH10</td>
</tr>
<tr>
<td>totCH11</td>
</tr>
<tr>
<td>totCH12</td>
</tr>
<tr>
<td>totCH2</td>
</tr>
<tr>
<td>totCH1</td>
</tr>
<tr>
<td>sumEval_Manage</td>
</tr>
<tr>
<td>totPath_Lab</td>
</tr>
<tr>
<td>totCH8</td>
</tr>
<tr>
<td>totCH6</td>
</tr>
<tr>
<td>totCH5</td>
</tr>
<tr>
<td>totCH3</td>
</tr>
<tr>
<td>two_years</td>
</tr>
<tr>
<td>start_2016</td>
</tr>
<tr>
<td>totMedicine</td>
</tr>
<tr>
<td>start_2015</td>
</tr>
<tr>
<td>wscore</td>
</tr>
<tr>
<td>one_year</td>
</tr>
<tr>
<td>three_years</td>
</tr>
<tr>
<td>non_msk_surg</td>
</tr>
<tr>
<td>start_2017</td>
</tr>
<tr>
<td>totcost</td>
</tr>
<tr>
<td>totCH17</td>
</tr>
<tr>
<td>less_one_year</td>
</tr>
<tr>
<td>totRad</td>
</tr>
<tr>
<td>start_2018</td>
</tr>
<tr>
<td>four_years</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>totAnesth</td>
</tr>
<tr>
<td>totCH15</td>
</tr>
<tr>
<td>totCH16</td>
</tr>
<tr>
<td>start_2014</td>
</tr>
<tr>
<td>totCH20</td>
</tr>
<tr>
<td>totdur</td>
</tr>
</tbody>
</table>
Multi-Class Neural Network Results ~ Predicting Escalation Category

**Metrics**

<table>
<thead>
<tr>
<th></th>
<th>Overall accuracy</th>
<th>Average accuracy</th>
<th>Micro-averaged precision</th>
<th>Macro-averaged precision</th>
<th>Micro-averaged recall</th>
<th>Macro-averaged recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI with...</td>
<td>0.64027</td>
<td>0.64027</td>
<td>0.64027</td>
<td>0.446537</td>
<td>0.446514</td>
<td></td>
</tr>
<tr>
<td>Multiple...</td>
<td>0.824513</td>
<td>0.824513</td>
<td>0.824513</td>
<td>0.446537</td>
<td>0.446514</td>
<td></td>
</tr>
<tr>
<td>No Escal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Confusion Matrix**

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>MRI with...</th>
<th>Multiple...</th>
<th>No Escal.</th>
<th>Surgery...</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI with...</td>
<td>9.3%</td>
<td>18.0%</td>
<td>47.6%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Multiple...</td>
<td>6.9%</td>
<td>55.0%</td>
<td>12.6%</td>
<td>25.6%</td>
</tr>
<tr>
<td>No Escal.</td>
<td>2.8%</td>
<td>2.5%</td>
<td>88.3%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Surgery...</td>
<td>8.3%</td>
<td>21.1%</td>
<td>45.0%</td>
<td>25.7%</td>
</tr>
</tbody>
</table>