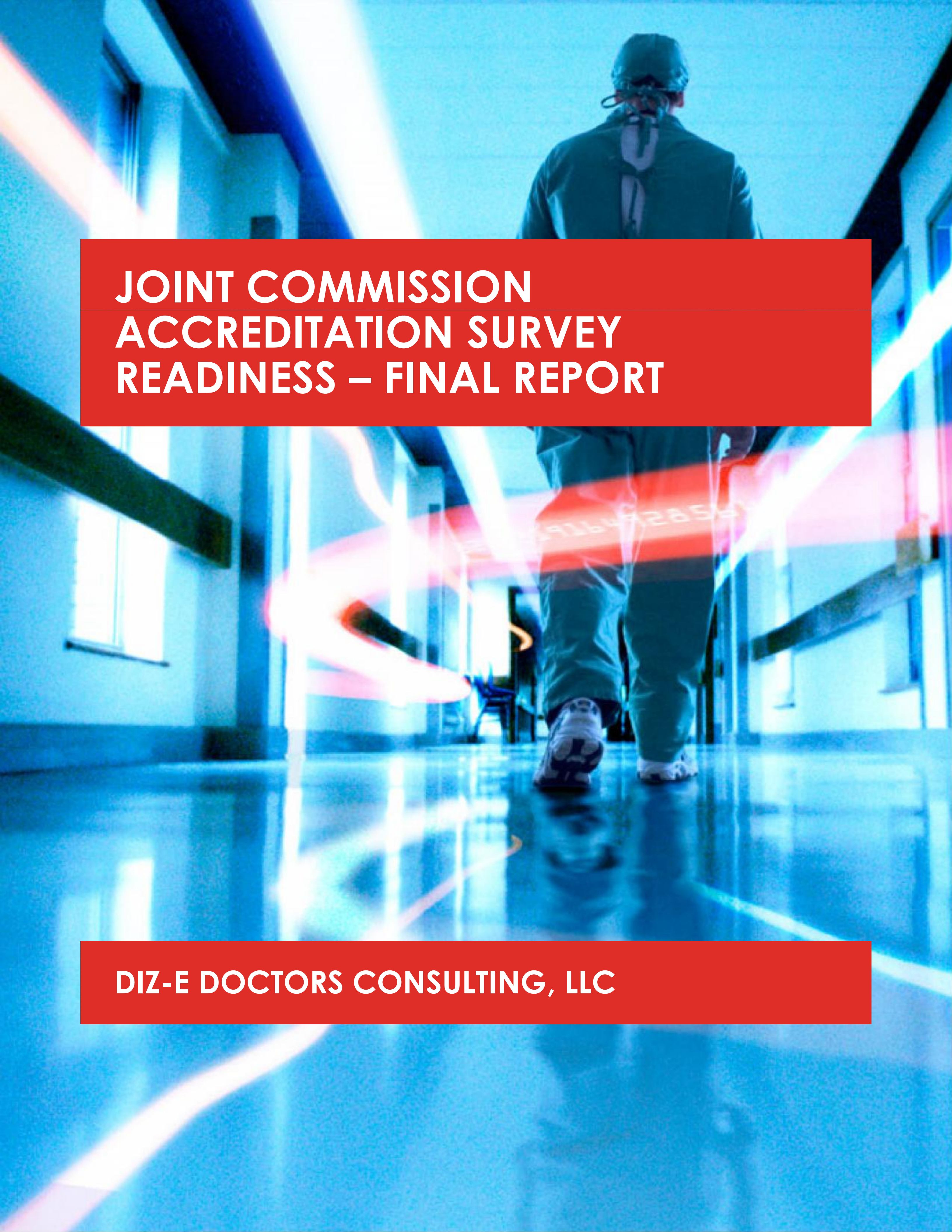


JOINT COMMISSION ACCREDITATION SURVEY READINESS – FINAL REPORT

DIZ-E DOCTORS CONSULTING, LLC

LETTER OF ENGAGEMENT

March 3, 2018

Dr. Donald Wedding, CEO & Minister of Health, Wellbeing and Mindfulness

Dear Dr. Wedding,

We respectfully submit for your review this initial finding document as per the agreed upon contract of work. The final report document includes the following:

- Definition of the problem statement
- A detailed data overview complete with data description, transformation steps and analysis process
- Conclusions including model variable selection and model results
- Key takeaways and recommendations from the analysis

We anxiously await your feedback and look forward to presenting our report in person to ensure it meets your expectations as a valuable resource for Joint Commission accreditation survey readiness and preparation. In the meantime, please feel free to contact any member of our team if you need additional information or materials for your Board of Directors or staff members.

Sincerely,

Diz-E Doctors Consulting (Team 04)

- Andrew Bland
- Eric Gero
- Richard Meyer
- Adam Sandstrom
- Tracy Valentine

CONTENTS

STATEMENT OF PROBLEM

Statement of Problem

INTRODUCTION

The number of deaths caused by patient safety issues in America's hospitals could fill a commercial jumbo jet each and every day. Imagine if the airline industry crashed a jumbo jet and killed that many people on a daily basis – there would no doubt be a public outcry for safety reforms on the part of the airline industry as a whole. Unfortunately the healthcare industry has made woefully insufficient progress in addressing what should be a top priority in all hospitals across the country. The Joint Commission (TJC) has an opportunity to engage institutions in focusing on high risk practices that will improve patient safety in America's hospitals and healthcare organizations.

TJC along with 3 other organizations accredits healthcare organizations as safe to conduct operations and receive payment from Medicare and Medicaid. As such, TJC has unique levers to drive a national patient safety agenda. TJC possesses an enviable position among accreditation organizations with over 78% market share. Unfortunately, like many organizations, TJC is data rich and information poor.

With our help TJC can tap into a large amount of unstructured data stored across multiple data sources and transform this into a structured program that will educate and enable hospitals to improve the care they deliver to patients. The impact we can have would be like preventing a fatal jumbo jet crash from occurring each and every day.

Imagine being able to stop a jumbo jet crash per day! This is the impact we can have on patient safety in the United States if we are able to help TJC improve patient safety. Since accreditation is the "seal of approval" that allows hospitals to treat and bill for Medicare and Medicaid patients, this is a golden opportunity to align patient safety, quality and financial incentives for a large share of the U.S. healthcare market.

PURPOSE OF THE PROJECT

We will transform the TJC's scattered and unstructured data into a meaningful display of risks faced by each hospital in the United States. This will allow TJC to redesign their survey process to focus on the "high-risk, high-yield" areas during the survey process that will most likely impact patient safety. This valuable resource can also be shared with the healthcare institutions whose financial solvency depends on a successful TJC survey.

STATEMENT OF PROBLEM

OBJECTIVES

- Illustrate the high risk, high dollar survey findings that impact patient safety to the survey team
- Develop models that predict the risk for future surveys
- Deliver a descriptive analytics visualization tool (dashboard) that combines electronic survey data with hospital demographic data including bed size, average daily census, number of care sites and visits to better understand the historical patterns and correlations between survey findings, severity score and hospital demographics.
- Develop predictive models that leverage the data analysis to assist hospitals across the United States in improving their quality and safety.

PROJECT GOALS & DELIVERABLES

- Establish a data integration methodology
- Correlation matrix analysis between all findings
- Heat map visualization of high severity findings
- Correlation matrix between high severity findings
- Heat map visualization of findings and demographic descriptions of each hospital
- Predictive model that assists hospitals in identifying their most probable high risk findings
- Dashboard illustrating survey findings risk and correlation for each type of hospital – "The Findings Finder"
- Interactive application for desktop or mobile device to present survey findings risk
- Provide valuable material for education and training of TJC employees and contracted consultants.

DATA OVERVIEW

Data Overview

Since changing to an electronic survey capture tool, TJC has captured tens of thousands of detailed accreditation survey event findings from over 4,000 U.S. hospitals and healthcare organizations from 2015–2017. TJC surveyors visit accredited health care organizations a minimum of once every 39 months (two years for laboratories) to evaluate standards compliance. This visit is called a survey.

DATA DESCRIPTION

- Compiled from three separate Joint Commission databases across three divisions from:
    - Electronic data capture tool which requires hospitals to self-report demographic information
    - Survey findings database that surveyors have used for the last 4 years to report findings electronically
    - Survey interpretation group data collection that summarizes the findings
- Originally sourced directly from surveys, public records and the hospital's CMS applications, providing a highly robust data set. Data was made available to the team in MS Excel file format.
- Data file facts (additional information available in the Appendix A)
    - Survey Findings – 126,726 total survey findings capturing data from 19 separate variables
    - Survey Event – 16,609 rows and 16 columns regarding the detailed TJC demographic information regarding the different volume measurements across each hospital and its different departments
    - Main Site/Secondary Site Characteristics – 4,701 rows of individual hospitals reporting 16 separate demographic variables
- Results of surveys are represented by findings and *Survey Analysis for Evaluating Risk* (SAFER) scores. SAFER scores were put into practice in 2017 and are not available on results data from prior years. Thus survey data can only be analyzed and modeled for severity with SAFER scored results.
- Data without SAFER scores will be analyzed and modeled for pairs or clusters of findings likelihood by hospital characteristics.
- No variables were removed since they can be utilized to tell a story about the data and/or modeling.
- Exploratory data analysis (EDA) was conducted in MS Excel, R, MySQL and Tableau.

DATA OVERVIEW

PROJECT DATA STORAGE HANDLING/ACCESS

Data storage handling, analysis and interaction tasks are handled by a variety of tools including a cloud-based data storage solution.

- AWS MySQL DB solution – a relational database hosted by Amazon's AWS cloud solution, currently on the free-usage tier and fully scalable to meet any future scope expansion
- MySQL Workbench – a freely downloadable database development client interface for data access, manipulation and administration tasks on the MySQL DB platform
- Research and analysis tools including Weka, R and Tableau, providing 24/7/365 direct data access to the MySQL database cloud storage

OVERVIEW OF THE DATA

The survey findings data contains some duplicated records that will be removed prior to analysis (Figure **1** below). The remaining data is segregated into survey results with findings only (pre-2017) and survey results with both findings and SAFER scores (2017 forward).

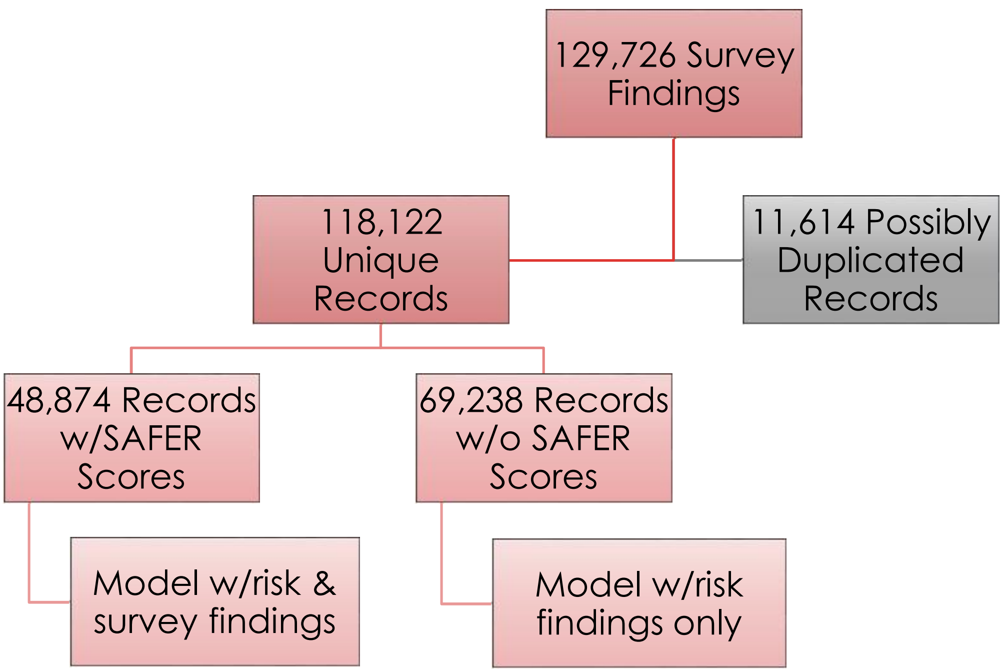


Figure 1: Survey findings data characteristics, illustrating how data was handled for analysis and modeling purposes

The R statistical analysis application was used to generate Figure 2, showing how the data are distributed across the 18 chapters of survey findings as defined by the TJC manual. The most frequent findings are in Environment of Care and Infection Prevention and Control. These 8,000+ findings dwarf the other chapters and present the most significant risk to the organizations undergoing survey.

DATA OVERVIEW

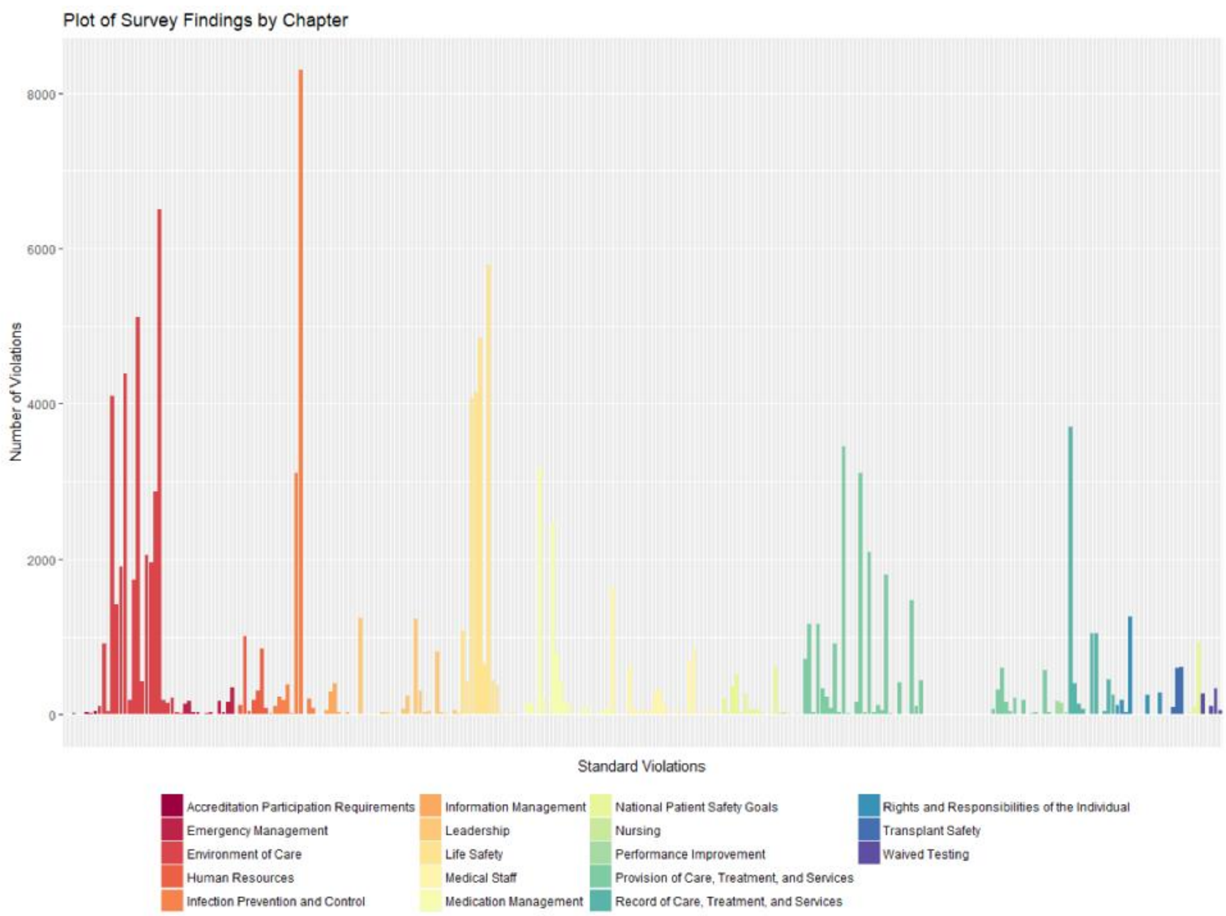


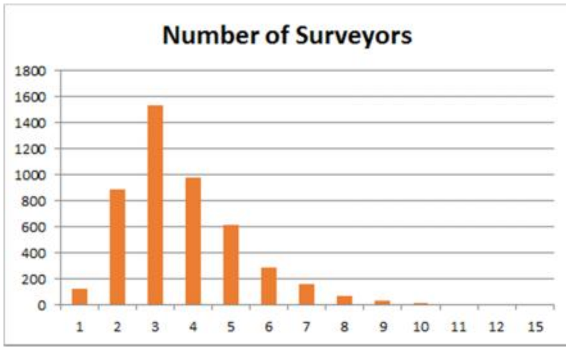
Figure 2: Number of survey findings by findings chapter

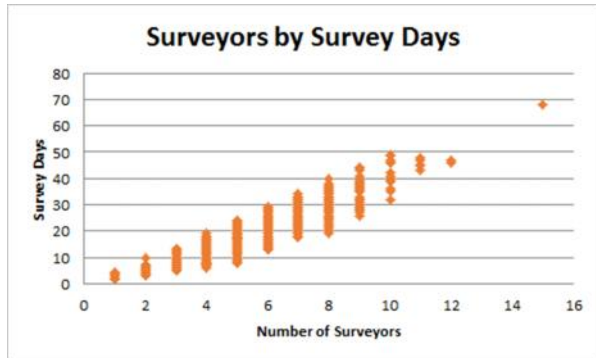
Figures **3** and **4** below, and Table **1** to the left show survey events by number of surveyors, and survey days by number of surveyors. The majority of surveys have 3 surveyors, but as many as 15 have been used in one instance.

Value	Total Surveyor Days	Number of Surveyors	Days per Surveyor
Minimum	2.00	1.00	1.50
Average	11.22	3.70	2.88
Maximum	68.00	15.00	5.00

Table 1: Surveyor days, number of surveyors and days per surveyor – range and mean values

DATA OVERVIEW





Figures 3 & 4: Survey events by number of surveyors, and survey days by number of surveyors

Although the overall number increases as more surveyors are added, there doesn't appear to be a visual correlation between more surveyors and more findings per surveyor in Figure **5**.

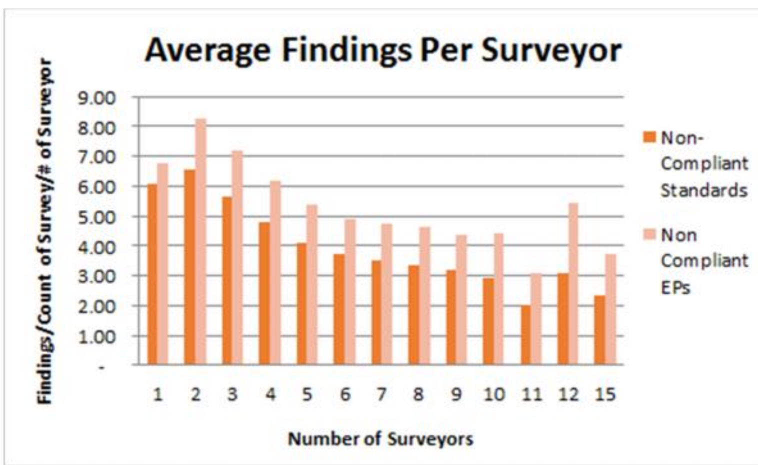


Figure 5: Average findings per surveyor, faceted by non-compliant standards and non-compliant EPs

The majority (95.7%) of the surveys are Unannounced Full Event surveys as seen in Table **2** below. The rest are Initial Unannounced Full Events with only 0.15% of all surveys occurring in the unannounced extension new program category. The vast majority of the survey decisions (86.43%) are pending, with the next largest being accreditation as seen in Table **3**.

DATA OVERVIEW

Survey Event Type	% of Surveys
Initial Unannounced Full Event	4.15%
Unannounced Extension New Program	0.15%
Unannounced Full Event	95.70%

Survey Decision	% of Surveys
Pending	86.43%
Accreditation with Follow-up Survey	10.74%
Preliminary Denial of Accreditation	2.60%
Accreditation with Full Standards Compliance	0.09%
Accreditation Denied	0.09%
Adverse Decision, Pending	0.06%

Tables **2 & 3**: Percent of surveys by survey event type and survey decision

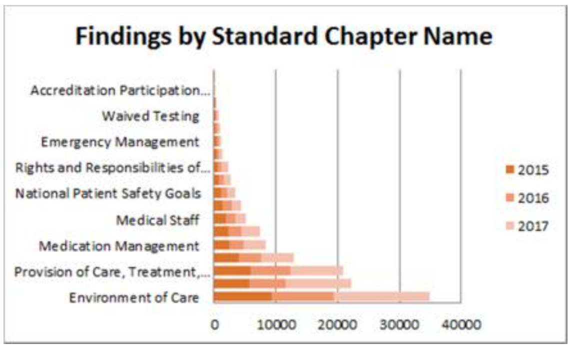


Figure **6 & 7**: Findings by Standard Chapter name, and SAFER score counts by Standard Chapter name

Environment of Care makes up for approximately 27% of the overall findings (Figure 6). This is followed closely by Life Safety, Provision of Care, Treatment and Services and Infection Prevention and Control. A SAFER score of 10 represents an immediate threat to life. Environment of Care is the most prevalent finding and makes up the majority of high SAFER scores (Figure **7**). Coming in a close second is leadership - when leadership breaks down, many pieces of the hospital suffer as a result.

Table **4** shows the SAFER score likelihood distribution of scores by standard chapter name, where SAFER scores are available. The likelihood scores are grouped by Low, Moderate, High and Immediate Threat to Life (ITL). As can be seen from this table, the accreditation participation requirements contain a disproportionate amount of immediate threat to life points than any other standard. Leadership and National Patient Safety Goals also have above average high likelihood values.

Standard Chapter Name	Low	Moderate	High	ITL
Accreditation Participation Requirements	37%	0%	2%	62%
Emergency Management	71%	18%	0%	0%
Environment of Care	59%	31%	7%	0%
Human Resources	4%	3%	1%	0%
Infection Prevention and Control	8%	12%	3%	0%
Information Management	1%	0%	0%	0%
Leadership	2%	3%	1%	0%
Life Safety	26%	5%	0%	0%
Medical Staff	3%	2%	0%	0%
Medication Management	3%	5%	1%	0%
National Patient Safety Goals	1%	2%	1%	0%
Nursing	0%	0%	0%	0%
Performance Improvement	0%	0%	0%	0%
Provision of Care, Treatment, and Services	8%	9%	1%	0%
Record of Care, Treatment, and Services	4%	2%	0%	0%
Rights and Responsibilities of the Individual	1%	1%	0%	0%
Transplant Safety	1%	0%	0%	0%
Waived Testing	0%	0%	0%	0%
Grand Total	**29%**	**18%**	**3%**	**0%**

Table **4: SAFER score likelihood distribution by standard chapter name**

Most secondary sites are within five miles of the main site as can be seen in Figure **8** below. This exhibits the count of secondary sites that fall within the mile range referenced.

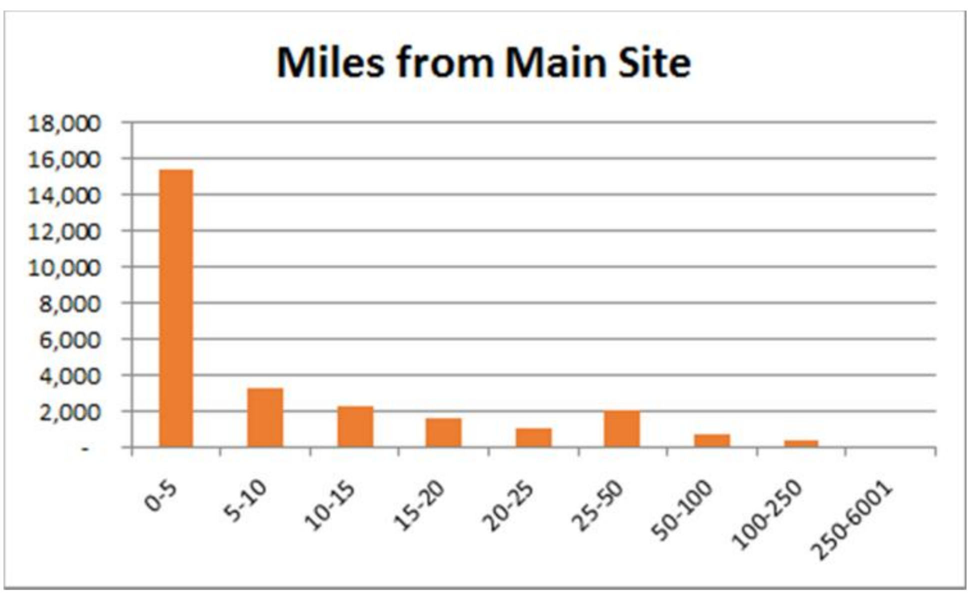


Figure **8: Miles that Secondary Sites are away from the Main Site**

After adding a variable for how full a hospital was, a histogram was created to exhibit the distribution based on main site classification. This new variable takes the average daily census divided by the available beds from the HAP data site. It is apparent in Figure **9** below that surgical hospitals operate at a lower capacity than Academic or Psychiatric.

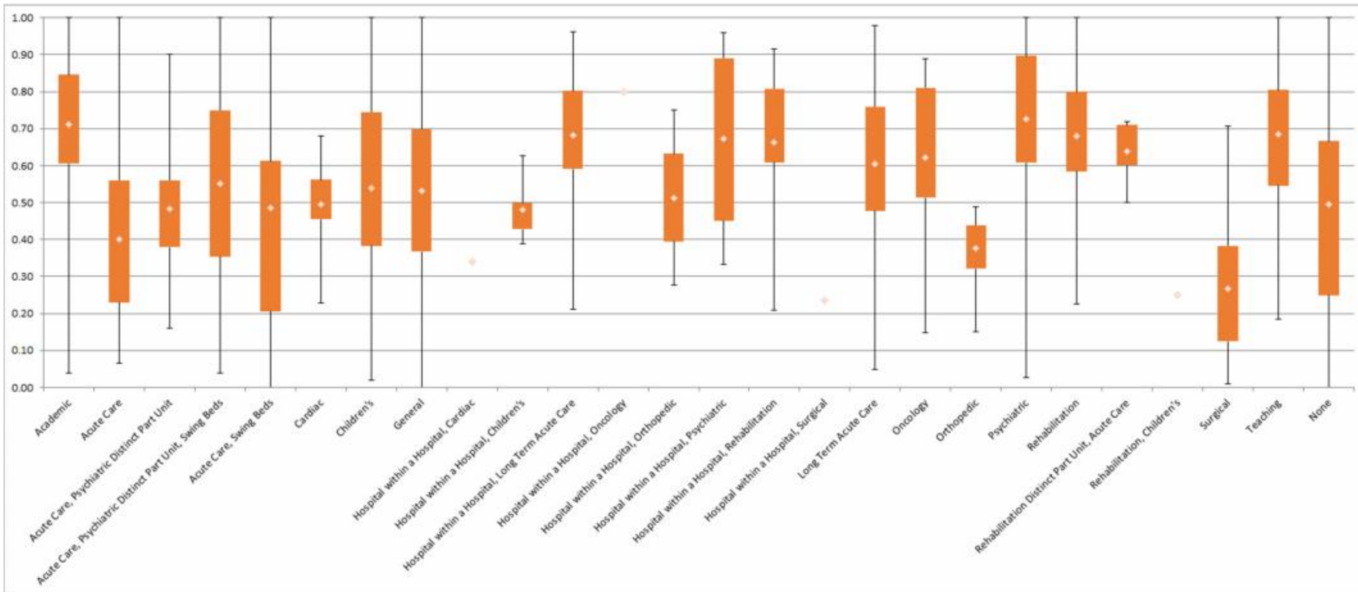


Figure 9: Box and Whisker plot showing distribution of Hospital Capacity

DATA TRANSFORMATION

Overall, the data integrity was acceptable and very few records contained missing values. Missing values for the main site variable in the Survey Findings and Survey Event tables were corrected by joining the secondary site ID to the Secondary Sites table to retrieve the missing values. The categorical variables were checked by reviewing their distinct values to ensure that all values adhered to the defined categories.

The continuous variables consisted of bed counts, census data, count of visits, count of surveyors, count of survey days, count of non-compliant standards, and the count of non-compliant elements of performance. Many of these variables are summaries of variables found in other tables, so it was possible to verify their accuracy directly to ensure that no outliers existed in the data.

Since the level of each primary table represents a different granularity, a great deal of time was spent combining the tables into one master table that could be used for modeling. Currently, several variations of the master table exist, each being the result of different variable combinations and variable roll-ups to support different modeling approaches for different response variables.

Several new variables have been created, mostly as the result of rolling up very granular measures to join with the main sites. The engineered variables include binned organization categories and settings, binned variables for bed size, average daily census, visit counts, and a utilization variable that is obtained by dividing the average census by the number of beds since it is hypothesized that organizations with high utilization are susceptible to higher frequencies of findings.

Since the data resides in a MySQL instance on Amazon RDS, most of the wrangling was done with SQL. Using SQL allows for the manipulation of large data sets with very little effort. In addition, once finalized, the SQL statements have been embedded into R scripts which not only keeps them organized, but also allows them to be run when needed to rebuild tables.

DATA ANALYSIS APPROACH

During the initial exploratory analysis, it was noted that of the approximately 122K survey findings reported, 271 unique standard findings and 1,421 unique elements of performance (EP) findings were observed. Since these findings represent a portion of the total standard and EP findings available it suggested that particular patterns of standard and EP findings are present. To better understand how each standard finding is related to other standard findings, the frequency of unique standard pairs by hospital type was determined. This analysis led to some interesting observations, since it shows that certain standard findings are more likely for particular hospital types, while other standard combinations rarely occur, if ever at all.

Such information is valuable to hospitals and healthcare facilities for performing a “what-if” analysis while preparing for a survey to determine the likelihood of incurring a particular standard finding based on another. This information will be made available to organization administration through a desktop/mobile application. It is also valuable to TJC since the presence of a standard pair finding with a low probability may be considered a red flag indicating additional problems requiring further investigation or the need for additional training of the surveyors. The analysis was also performed for EPs and standard and EP labels, with similar results.

Figure **10** below shows examples of standard pair probabilities by hospital type. The plot on the top left shows that an Acute Care, Psychiatric Distinct Part Unit has a 90% chance of incurring a Provision of Care, Treatment, and Services - PC.01.02.13 finding with an Infection Prevention and Control - IC.02.02.01 finding and that several other findings are likely as well. The plot on the bottom right shows that an Acute Care, Swing Beds facility has a low likelihood of incurring any particular finding with a Life Safety - LS.01.01.20 finding.

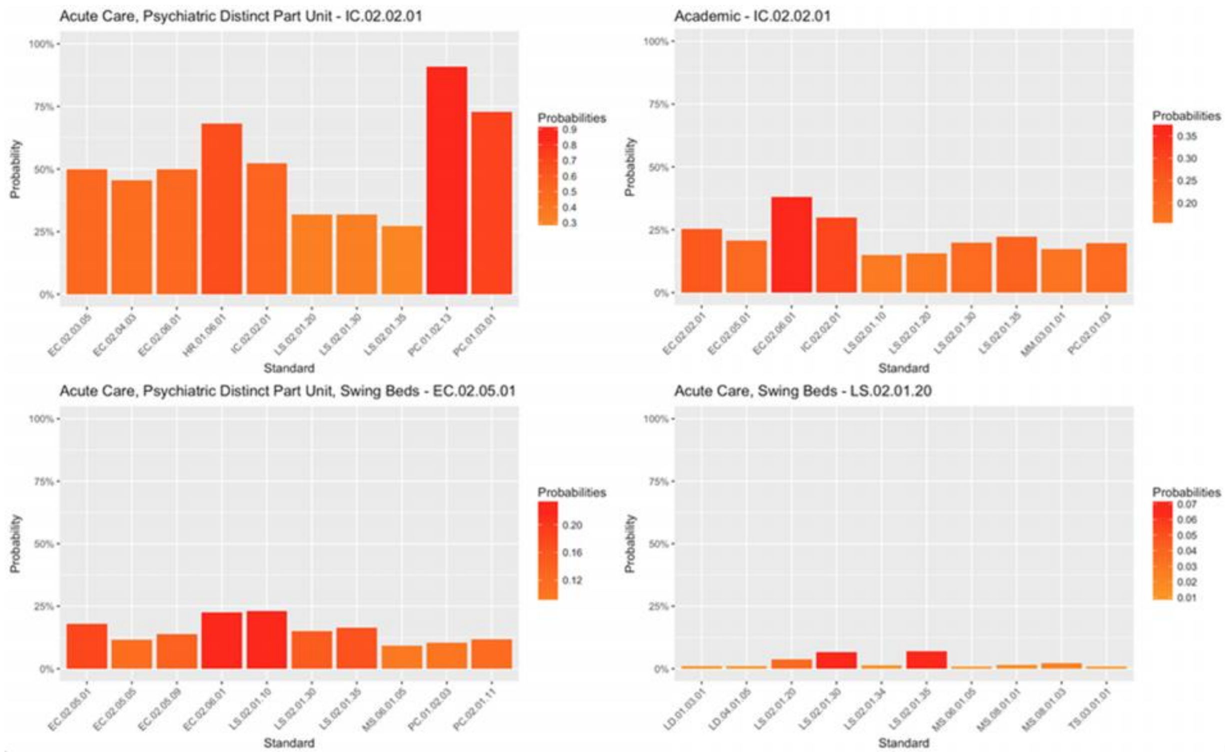


Figure 10: Examples of standard pair probabilities by hospital type

Conclusions

VALUE PROPOSITION

In an ideal world, healthcare organizations would operate at all times as though they are being evaluated by The Joint Commission. The study titled *Patient Mortality During Unannounced Accreditation Surveys at US Hospitals* by Michael L. Barnett MD, Andrew R Olenski BS, and Anupam B. Jena MD PhD suggests that this may not be the case. The study shows that mortality rate of patients admitted during an Unannounced Accreditation Survey was lower than that as compared to patients admitted three weeks prior or post survey. These findings demonstrate that heightened attention to detail does improve outcomes, in this case mortality.

Our analysis, when shared with organizations has the potential to improve patient outcomes and save more lives by extending the brief window of performance improvement experienced during the survey even further towards the continuous end of the spectrum. The cluster analysis shows there is a non-intuitive grouping of survey findings that is not associate with obvious characteristics such as size, type of hospital, number of patients served, or number of surveyors on site during the survey.

An organization's state of preparedness relating to a survey is commonly referred to as readiness. Organizations often aspire to a state of continuous readiness but this is more difficult in reality. There are approximately 2,000 EPs which make it extremely challenging to achieve ongoing readiness. The efforts made to this point have been aimed at understanding patterns that exist in the findings. Understanding these patterns serve member organizations by helping them to prioritize ongoing readiness efforts. These patterns will also benefit The Joint Commission in the executions of survey events, surveyor training and evaluations, and survey effectiveness.

Further, future studies will be aided through the development of data processing and variable transformations established in the creation of these findings. Reducing findings into clusters may make it easier to understand how decisions made at the organizational level are likely to result in survey findings. Linking those findings to patient outcomes would then complete the loop providing value to healthcare organizations, The Joint Commission, and Centers for Medicare & Medicaid Services (CMS).

CLUSTER ANALYSIS OF SURVEY FINDINGS

Cluster analysis is the last portion of the exploratory data analysis. This technique is used when there are a large number of variables and the associations between those variables is not known or clear. In this case the association between the 2,000 elements of performance, hospital type, survey year and SAFER score was not previously known and too complex to be visualized.

The Joint Commission desired to understand the relationships between these variables and determine if these associations were valuable enough to report back to their hospitals as potential associations.

Currently The Joint Commission simply reports the "top 10 citations" for each type of hospital to all of their customers. Customers have requested a more robust association between the citations and hospital characteristics.

Some example questions we wish to answer with cluster analysis are:

a) For a General Hospital of my size and associated with a system, what are the findings that I need to pay most attention to during a survey and between surveys?
b) If I have a finding y, what findings x1, x2, x3, x4…..xn should I be looking for?
c) Can you predict all of the findings a hospital like mine will have?
d) Has the type of findings you are seeing changed in a hospital like mine since I was last surveyed?
e) How does my psychiatric hospital differ from my acute care hospital based on findings?
f) Does the number of surveyors you send to my hospital impact the number or severity of my findings?

To begin assessing these questions, an unsupervised cluster model was used to identify clusters. The initial model suggested that five clusters achieved optimal separation with small incremental improvements the more different clusters were included.

CONCLUSIONS

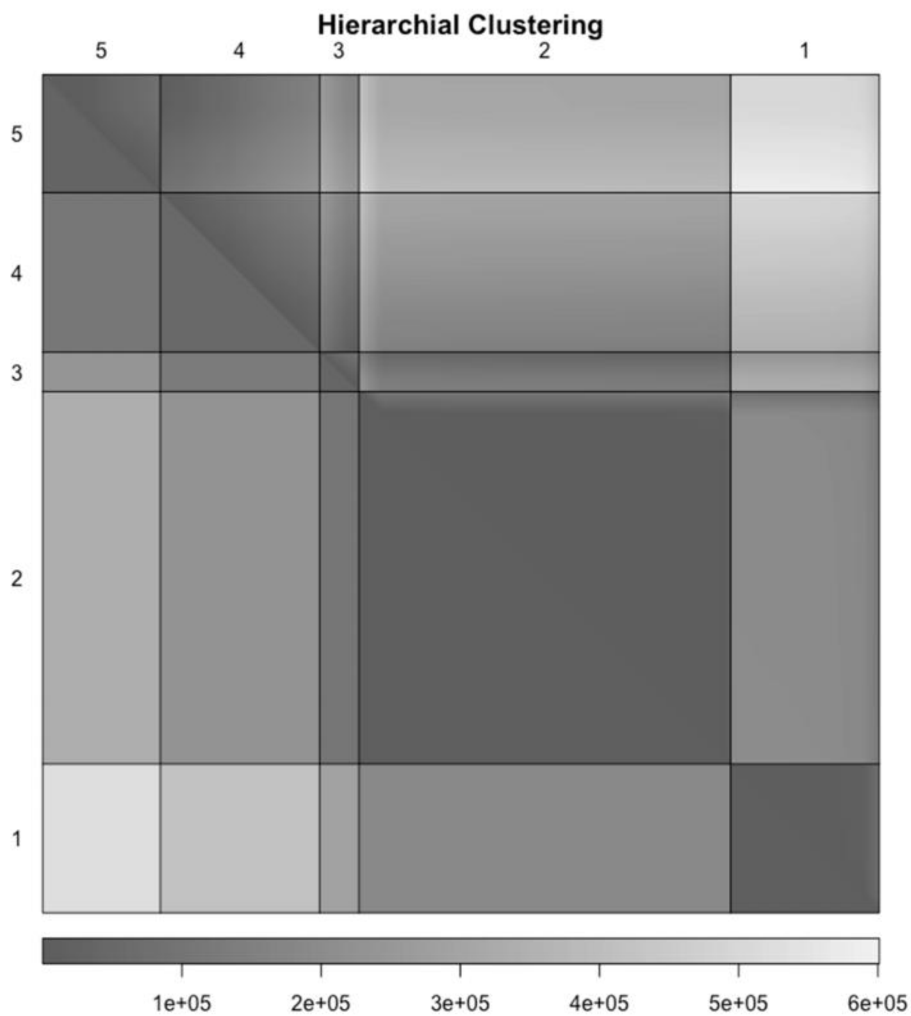


Figure 11: Confusion matrix and heat map illustrating results from clustering method

CONCLUSIONS

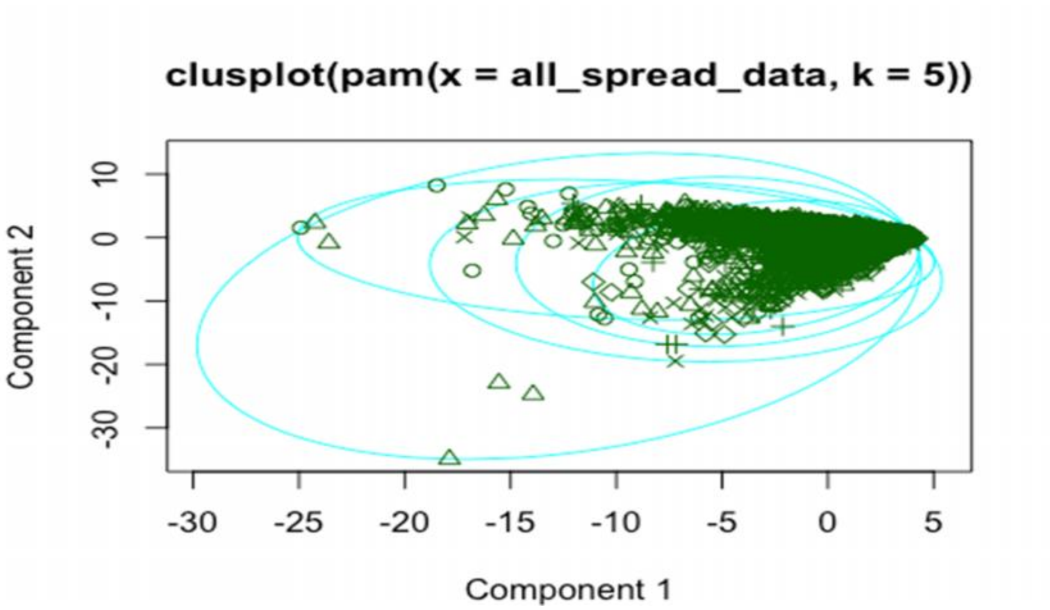


Figure 12: These two components explain 5.32% of the point variables

However, component analysis suggested the hospital site number played in inordinately important role in this clustering approach and the visualization of these different groups did not provide any additional information to The Joint Commission.

Reprocessing the data excluding the hospital number and including some additional demographic data showed clusters (k-mediods model) that made sense intuitively and assisted in answering the questions The Joint Commission was asking about the data.

CONCLUSIONS

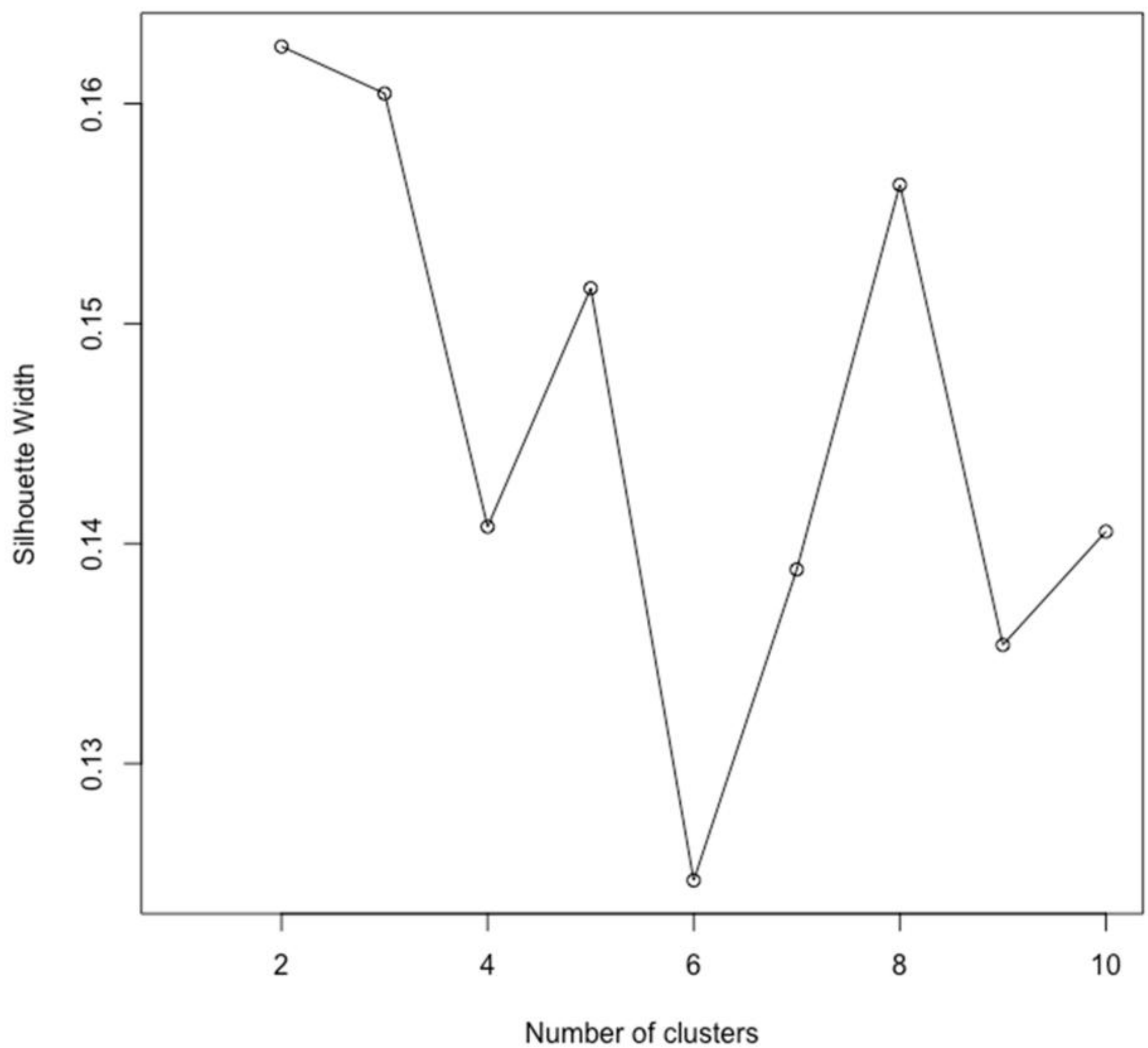


Figure 13: Outline showing optimal number of clusters

Figure **13** above suggests looking at the 2,3, and 8 cluster groupings to assess the value.

The 2 group cluster major differences was between surveys done in the 2015-2016 time frame before SAFER scoring and 2016-2017 time frame after SAFER scoring was implemented showing there was a significant difference between findings before SAFER and after.

CONCLUSIONS

The three cluster survey method showed a difference between 2015, 2016 and 2017 with each group containing hospitals only from one year. This again suggests that something different was occurring with the surveys during the Pre-SAFER year (2015) the transition year (2016) and the Post-SAFER year 2017.

The eight cluster group was the next group that contained the most similar clustering.

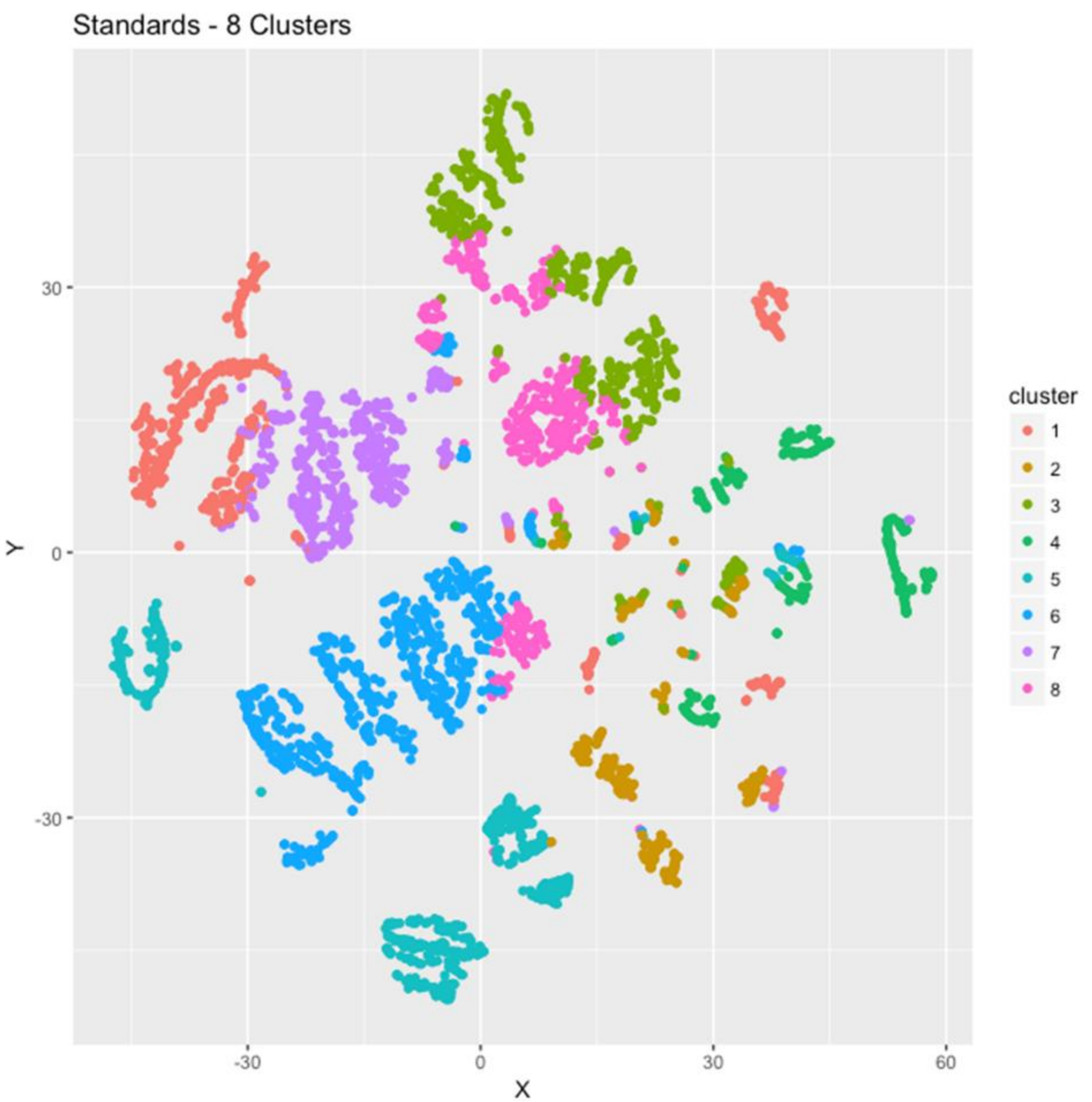


Figure 14: Groupings of Eight Cluster Analysis

CONCLUSIONS

Two clusters 1 and 3 had a grouping of hospitals based on the 2015 and 2016 survey years. The remaining 6 clusters began separating the hospitals into hospital type based on the characteristics including the pattern of survey type and findings.

This suggests the data to be used for modeling to predict findings, SAFER score and hospital type be from the year when SAFER scores were measured (2016-2017).

Going forward into the modeling, the subgroup selected will have a SAFER score greater than one to assist in predicting hospital type and survey findings.

Some example answers to the question above based on the cluster analysis:

a) For a General Hospital of my size and associated with a system, what are the findings that I need to pay most attention to during a survey and between surveys?
    a. Modeling using data with SAFER scores will be used.
b) If I have a finding y, what findings x1, x2, x3, x4…..xn should I be looking for?
    a. Modeling using SAFER score data will be used as well as a display dashboard and mobile application.
c) Can you predict all of the findings a hospital like mine will have?
    a. Yes, but with data after 2016 because the 2015 data does not look like 2017.
d) Has the type of findings you are seeing changed in a hospital like mine since I was last surveyed?
    a. Yes
e) How does my psychiatric hospital differ from my acute care hospital based on findings?
    a. It differs by findings and other demographic characteristics and the data suggests that models can detect and describe those differences.
f) Does the number of surveyors you send to my hospital impact the number or severity of my findings?
    a. Not in a major way when compared to class of hospital.

MODEL VARIABLE SELECTION

The attribute selection tool found in Weka was used to identify potentially important variables through multiple decision trees. There is consistency in the variables that were selected when predicting SAFER Scores, and standard or EP findings. Bed count, average daily census, and visits count which are among the few continuous variables present were not selected in any of the models. This came as a bit of a surprise since these variables relate directly to the size and complexity of the organization.

The variables hospital type and number of program sites were selected as top contenders in all models. Other variables that were selected often include standard label, count of Standard Chapter Name, Total Surveyor Days and the Count of Surveyors.

CONCLUSIONS

MODEL RESULTS

Aside from clustering, two main focus areas were selected for predictive modeling. The first was to predict the hospital type based on the standard findings. While this may seem counterintuitive since hospital type is known and established by an organization prior to the survey, determining if enough information existed in the data to accurately predict hospital type would ensure that healthcare organizations do indeed share identifiable commonalities. Such information would benefit The Joint Commission and healthcare organizations by allowing for targeted auditing and process improvements.

The second area of focus was to predict the SAFER score that an organization was likely to incur based on hospital type and standard findings. Recall that a severe finding could result in a SAFER score of 10 which could cause an organization to lose accreditation. The goal is to predict the likely range of SAFER scores that an organization is likely to receive based on previous surveys.

Several modeling methods were attempted, including linear models, decision trees, Gradient Boosting models, and Random Forests for regression and Support Vector Machines [SVM], Generalized Linear Models [GLM], decision trees, Gradient Boosting models, and Random Forests for multiclass classification. With 26 distinct levels for the hospital type variable, the performance and reliability of the SVM and GLM methods for multiclass classification were less than satisfactory and will not be discussed further.

Regression

The first model was a linear regression model to predict the SAFER score. A second model was then attempted to predict the maximum SAFER score after rolling up the data to the distinct standard and hospital type level. The Residuals vs Fitted plot in Figure **15** demonstrates the issues that were encountered with simple linear regression, likely caused by the multiple levels of the categorical variables Standards and hospital type as a predictors. A decision tree using the M5P algorithm was then fitted with similarly poor results.

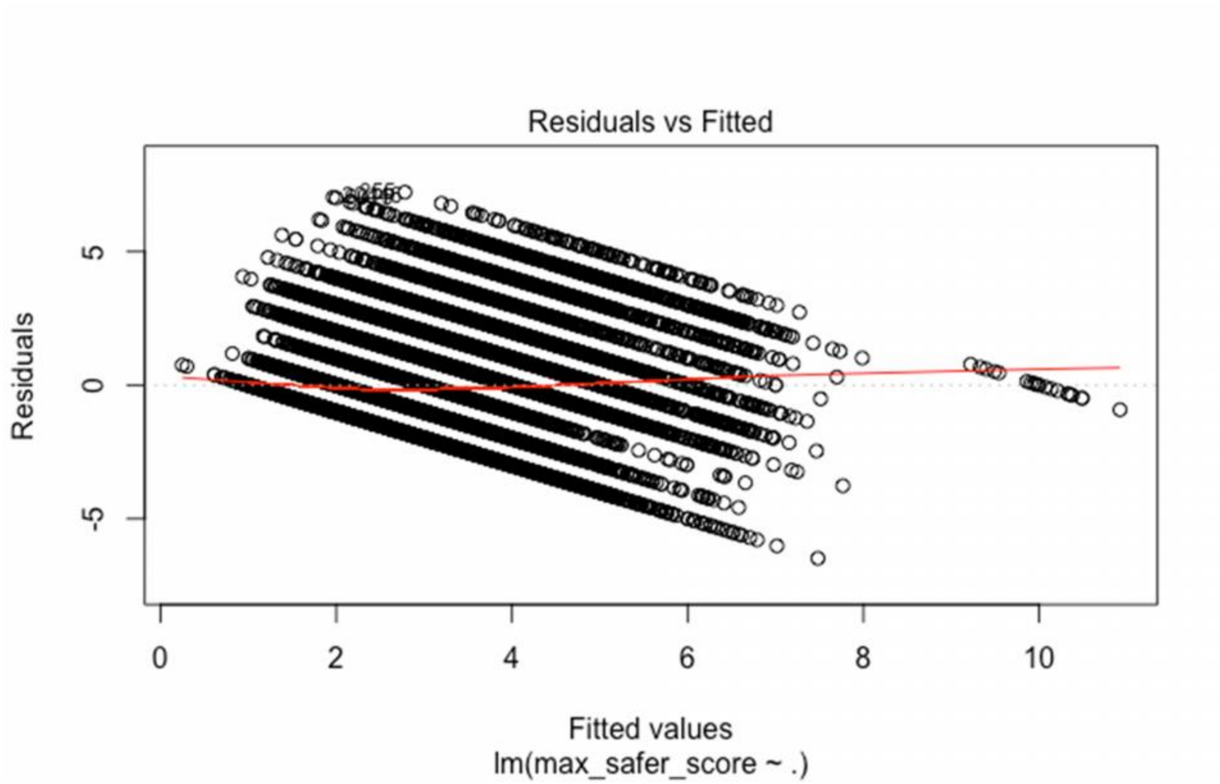


Figure 15: Residuals vs Fitted for Simple Linear Model

The results from the gradient boost and random forest models were not much better, but overall the best models were able to achieve a root mean square error [RMSE] of ~1.9. While this is less than ideal, it is acceptable for this initial round of providing an adequate SAFER Score range prediction. Ironically, the linear regression model had the lowest RMSE of 1.86. Table **5** below shows the evaluation metrics for regression models.

Regression					
Model	**Response Variable**	**Package**	**RMSE**	**Adjusted Rsquared**	**MAE**
Linear Regression	SAFER Score	stats	1.8600	0.2488	
XGBoost	Max SAFER Score	xgboost	1.8901		1.5529
Gradient Boost	Max SAFER Score	Caret	1.9072	0.2289	1.5958
Gradient Boost - Tuned	Max SAFER Score	Caret	1.8754	0.2449	1.5173
M5P	Max SAFER Score	RWeka	1.8802	0.2444	1.5249

Table 5: Regression Model Summaries

CONCLUSIONS

Classification

Several models were attempted to predict the maximum SAFER Score as a nominal variable, but were not very successful. The best, an X Gradient Boosting model, had an Accuracy of 0.3990 and a Kappa of 0.6874. Predicting hospital type also posed some initial challenges. The first model was a Random Forest model that contained only the standard label and hospital type variables. While there were few surprises as to which standards were considered most important in the model, the model was only able to predict with 60% accuracy and a Kappa of 0.0912. The variable importance plot is provided in Figure **16**. A similar X Gradient Boosting model provided nearly identical results.

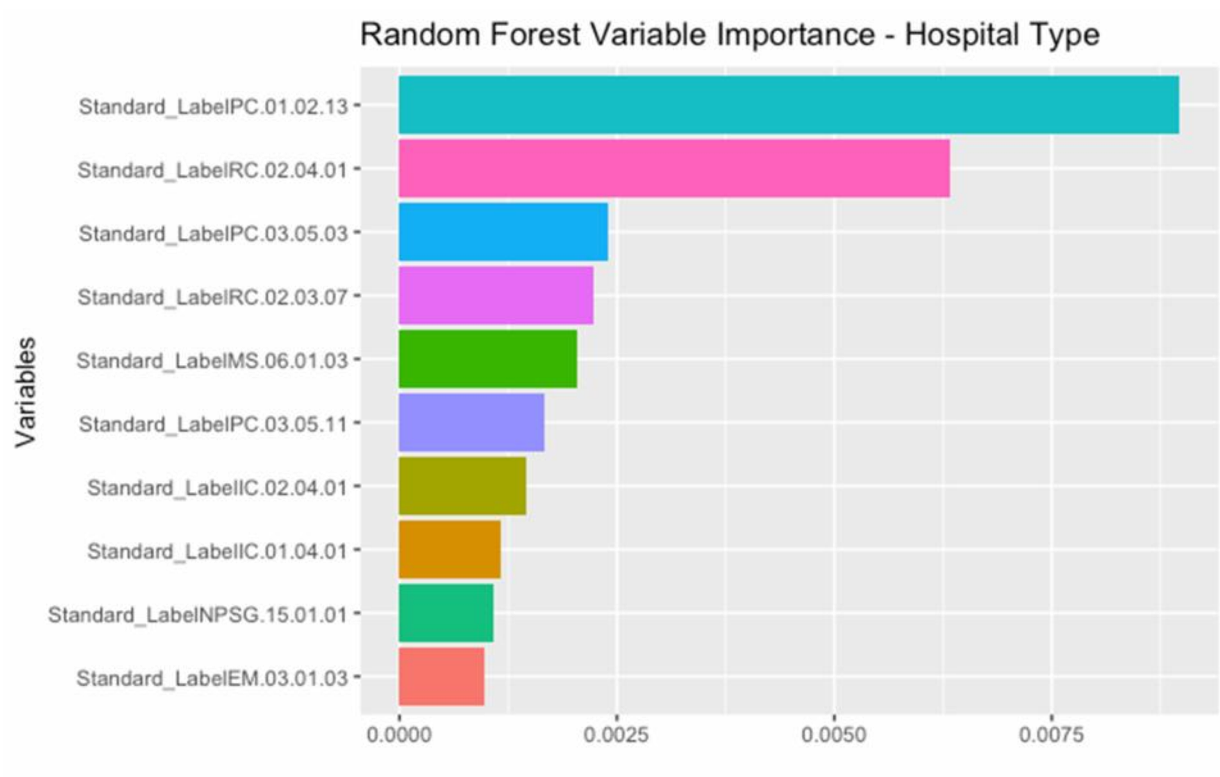


Figure **16: Random Forest Model Variable Importance – Hospital Type**

Adding variables to the model proved to be much more successful with the Random Forest and X Gradient Boosting models achieving an accuracy of 99%. Gradient Boosting models work by building weak learners (decision trees), then evaluating a loss function. The model essentially improves itself over multiple iterations each time it adds a new weak learner that minimizes the loss. The algorithm stops once it reaches a point where no new trees can be found or added that minimize the loss function being evaluated (Brownlee, 2016).

CONCLUSIONS

Figure **17** is a variable importance plot from the X Gradient Boosting model. The number of program sites for the organization played a large role in predicting hospital type, with the total number of surveyor days, and the survey year contributing as well. This is not at all what was expected, since we were hoping and expecting that the standards would be the primary variable in predicting hospital type.

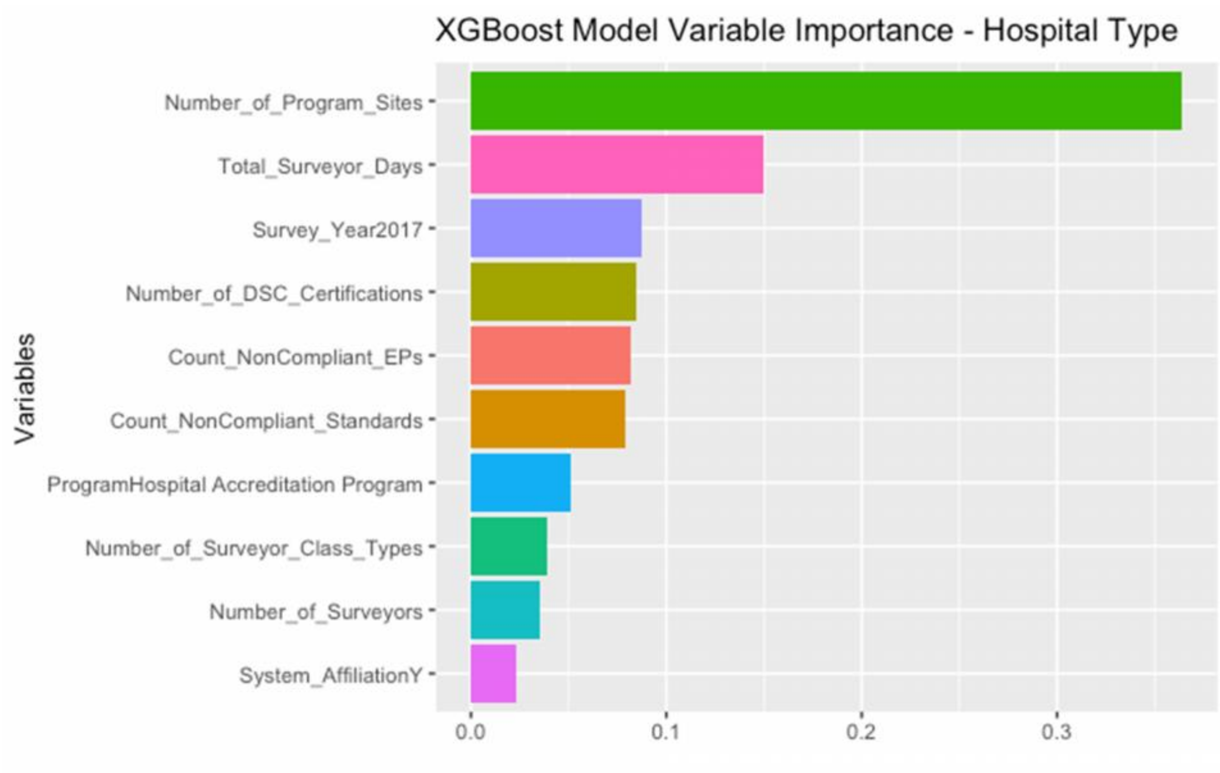


Figure **17: XGBoost Model Variable Importance – Hospital Type**

Figure **18** is a variable importance plot from the Random Forest model. This model had similar performance to the X Gradient Boosting model. Note how the number of program sites and total surveyor days are the top two predictors, but all other predictors vary in importance from that of the X Gradient Boosting model. The Random Forest model works by building multiple independent decision trees attempting to remove a defined bias.

The near perfect predictions on the test set by two separate modeling methods ensure that the data does contain enough information to distinguish hospital types. Also, since the data is at the Standards level, the prediction of the hospital type is a valid way to back into the Standard and hospital type pairs to understand which hospital types are associated with particular Standards.

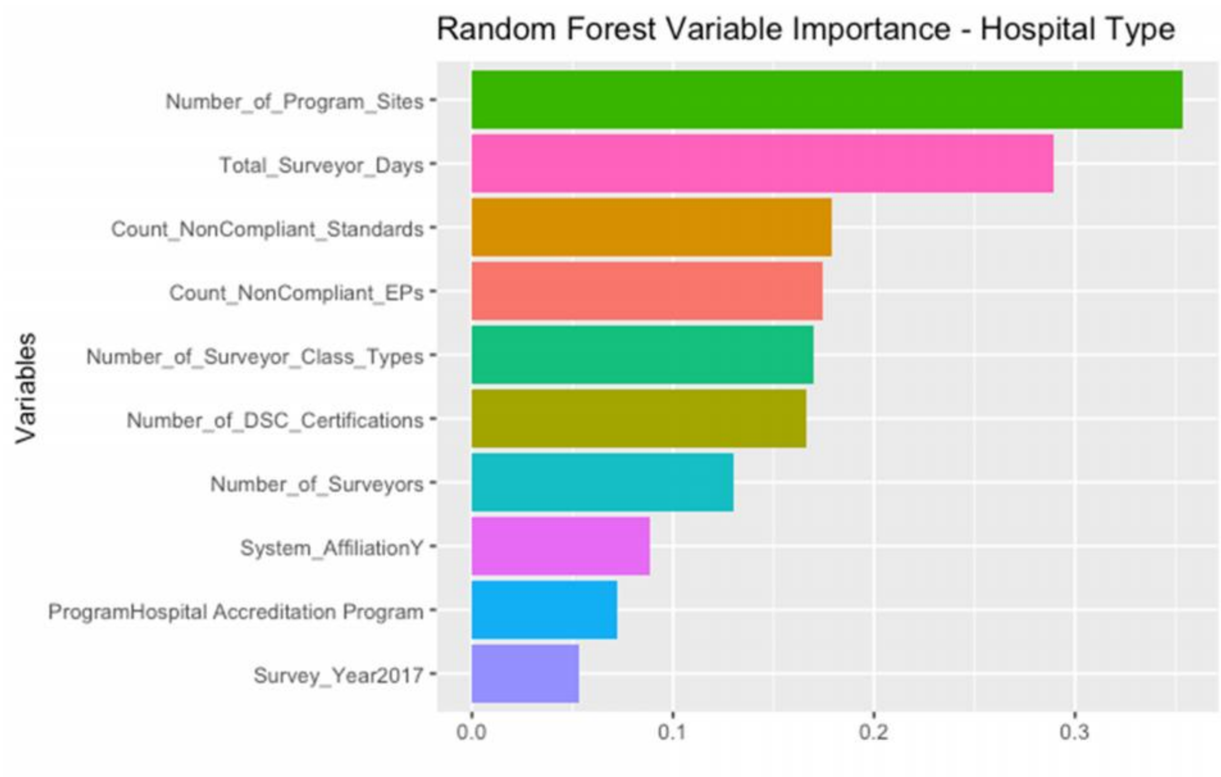


Figure **18: Random Forest Variable Importance – Hospital Type**

Table **6** provides a table of evaluation metrics for the classification models. Note how the different models for each response variable provided similar results.

Classification				
Model	**Response Variable**	**Package**	**Accuracy**	**Kappa**
Random Forest	Hospital Type	Caret	0.9906	0.9846
XGBoost	Hospital Type	xgboost	0.9927	0.9880
Gradient Boost	Hospital Type	gbm	0.8215	0.6874
Gradient Boost	Max SAFER Score	Caret	0.3897	0.1438
Gradient Boost	Max SAFER Score	gbm	0.3926	0.1533
XGBoost	Max SAFER Score	xgboost	0.3990	0.1574
J48	Max SAFER Score	Rweka	0.3326	0.1238
Random Forest	Hospital Type	Caret	0.6061	0.0912

Table 6: Classification Model Summaries

CONCLUSIONS

Many of these models are very resource intensive, taking a long time to run and resulting in large objects that needed to be cleared before additional models could be fitted. One way around this was to serialize the model objects ("saveRDS") and save them to the computer. This allowed us to reload ("readRDS") the models when needed to compare results or get predictions without needing to run the models again. This time saving method was shared by the team on the Information Sharing thread in Canvas for other teams to benefit from. Another method that was used to save time was parallel processing. This was from a tip shared by Team 3. It is not available for all modeling methods, but it does significantly improve the run time for Random Forests through the 'caret' package.

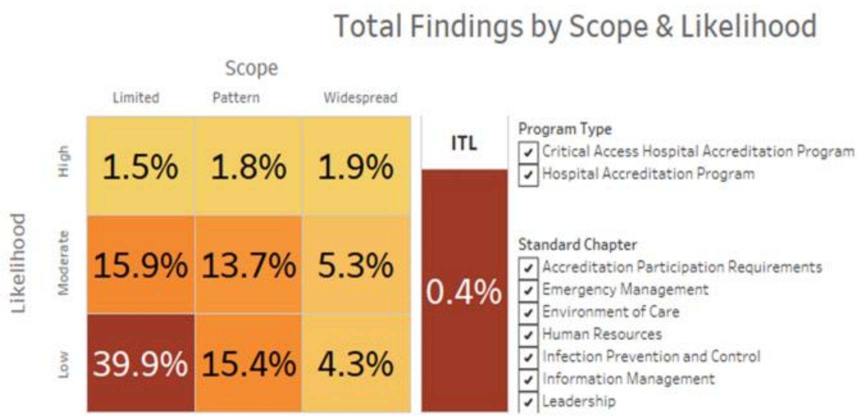
TAKEAWAYS

The existence of clusters that are not differentiated by demographics but by findings suggest that something else is creating these outcomes. Possible factors could be geographic location, individual surveyor(s) responsible for survey, hospital staffing mix, or countless other possibilities. Additionally, there is some concern regarding the granularity of the data. This is particularly true of the standards and EPs. Each of the roughly 270 standards belongs to one of 18 standard chapter names. Models built on standard chapter names do not appear to have enough information to pick up on the importance of the standards, while models built using the individual standards treat each standard as an individual entity. A middle level between standard and standard chapter name that groups standards by their similarities may provide the right level of detail to use standards as a useful and accurate predictor.

DASHBOARD OVERVIEW

A descriptive analytics visualization tool (a.k.a. dashboard) was used as an important tool for project data analysis. The dashboard combined electronic survey findings data with hospital demographic data including bed size, average daily census, number of care sites and visits to provide a better understanding of the historical patterns and correlations between survey findings, severity score and hospital demographics. The final dashboard will be provided as a tool for similar use by TJC decision makers and optionally the hospital systems they at which they perform surveys. Leveraging such a visualization tool should enable quick analysis and effective presentation of trends to important stakeholders and clients of TJC.

CONCLUSIONS



Program Type	1	2	3	4	5	6	7	8	9	10
Critical Access Hospital Accreditation Program	3.92%	1.38%	0.35%	1.47%	1.16%	0.52%	0.12%	0.13%	0.13%	0.04%
Hospital Accreditation Program	36.02%	14.00%	3.93%	14.46%	12.55%	4.78%	1.35%	1.62%	1.72%	0.36%

Standard Chapter Name	1	2	3	4	5	6	7	8	9	10
Accreditation Participation Requirements	0.01%	0.01%	0.01%				0.00%			0.06%
Emergency Management	0.48%	0.10%	0.06%	0.06%	0.05%	0.04%		0.00%	0.00%	
Environment of Care	10.99%	5.15%	1.53%	3.85%	3.62%	1.33%	0.28%	0.46%	0.71%	0.07%
Human Resources	0.86%	0.24%	0.06%	0.48%	0.33%	0.09%	0.08%	0.06%	0.05%	0.03%
Infection Prevention and Control	1.94%	0.87%	0.25%	1.41%	2.09%	0.87%	0.12%	0.33%	0.47%	0.08%
Information Management	0.36%	0.09%	0.03%	0.08%	0.07%	0.02%	0.01%	0.00%	0.00%	
Leadership	0.55%	0.31%	0.14%	0.39%	0.49%	0.31%	0.14%	0.21%	0.27%	0.10%
Life Safety	13.21%	3.53%	0.49%	1.59%	1.26%	0.21%	0.05%	0.06%	0.02%	
Medical Staff	1.21%	0.43%	0.24%	0.51%	0.46%	0.43%	0.03%	0.03%	0.04%	0.00%
Medication Management	1.67%	0.62%	0.13%	1.80%	1.29%	0.36%	0.17%	0.16%	0.04%	
National Patient Safety Goals	0.47%	0.12%	0.03%	0.88%	0.35%	0.16%	0.23%	0.13%	0.10%	0.04%
Nursing	0.02%	0.00%	0.01%	0.00%	0.00%		0.00%		0.00%	
Performance Improvement	0.07%	0.04%	0.03%	0.01%	0.05%	0.05%	0.00%	0.01%	0.01%	0.01%
Provision of Care, Treatment, and Services	4.27%	2.00%	0.65%	3.79%	2.71%	0.99%	0.29%	0.28%	0.09%	0.01%
Record of Care, Treatment, and Services	2.33%	1.07%	0.38%	0.63%	0.54%	0.28%	0.03%	0.02%	0.02%	0.00%
Rights and Responsibilities of the Individual	0.77%	0.39%	0.11%	0.29%	0.21%	0.05%	0.03%	0.01%	0.00%	0.00%
Transplant Safety	0.37%	0.29%	0.10%	0.04%	0.11%	0.05%	0.01%	0.01%		
Waived Testing	0.36%	0.11%	0.03%	0.12%	0.07%	0.04%		0.00%	0.00%	

Figure **19: View of Tableau dashboard developed to enable rapid understanding of complex survey data relationships and trends**

The dashboard depicted in Figures **19** and **20** show a survey SAFER score grid associated with the findings based on scope and likelihood of each finding's score. The scores range from 1 – 10, with 10 representing an Immediate Threat to Life (ITL) – the most severe score and could cause a hospital shut down to occur. More information on TJC's SAFER Matrix scoring can be found at the TJC website[1].

CONCLUSIONS

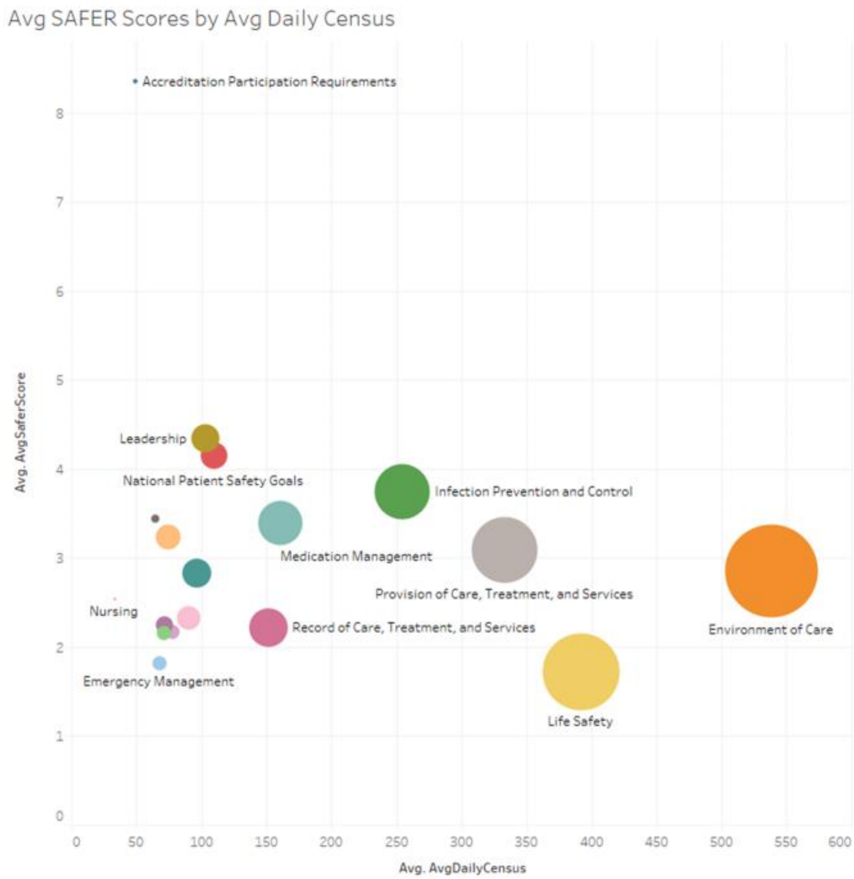


Figure **20: Another Tableau dashboard prototype view**

For presentation purposes, this dashboard was uploaded to the Tableau Public hosting site for general access. The TJC may wish to obtain their own enterprise Tableau server for internal/external dashboard hosting, with complete control over dashboard access privileges. Fellow course-mate Susan Chauhan's discussion forum posting Tableau Video Tutorial[2] was immensely helpful in the early stages of developing this dashboard.

The Dashboard is available at:

https://public.tableau.com/profile/rich3895#!/vizhome/PRED498_Team04_Project3/SurveyFindings-SAFERScoreGrid

MOBILE APPLICATION OVERVIEW

R Shiny and Amazon RDS were combined to develop an application that will be made available to TJC staff and healthcare organizations to assist in the preparation of survey events.

CONCLUSIONS

Figure **21** provides a view of the Standards tab. From this tab, the user can select a hospital type and a standard to view the minimum, average, and maximum SAFER scores predicted for that hospital type and standard pair. The plot shows the top five standards and their probabilities that are likely for the standard and hospital pair. From the screenshot below it is clear that for an Academic healthcare organization an Infection Prevention and Control - IC.02.02.01 is a reasonably serious finding and there is at least a 30% chance of incurring three of the five standards listed in the plot.

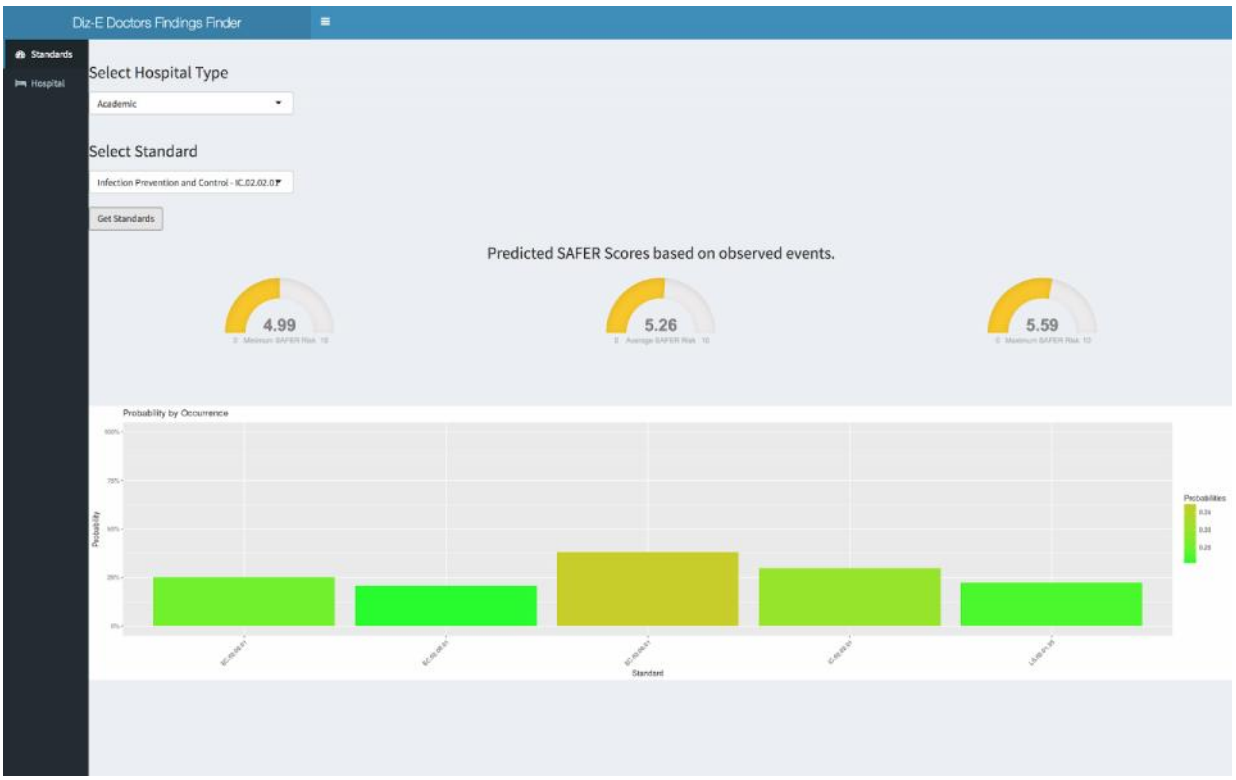


Figure **21: Screenshot of R Shiny Mobile Application SAFER Score Predictions**

Figure **22** provides a view of the Hospital tab. On this tab, the user can select a standard and retrieve a list of hospital types associated with that standard and the maximum predicted SAFERsScore. Since the number program sites was found to be a significant predictor for hospital type, an organization can look at the program sites that most closely matches the number of program sites they have and see which hospital type their organization most closely resembles. Ideally, an organization should most closely resemble their actual hospital type. If they resemble another type of organization that may indicate that they are reviewing a Standard that is not likely for their profile or their operations and procedures differ from other healthcare facilities of that type.

CONCLUSIONS

As previously discussed, survey events prior to 2017 were not SAFER score rated, so no information was available for some hospital and standard pairs. In such cases, the average SAFER score prediction for the standard was used. This will be corrected as more SAFER scores become available from current surveys.

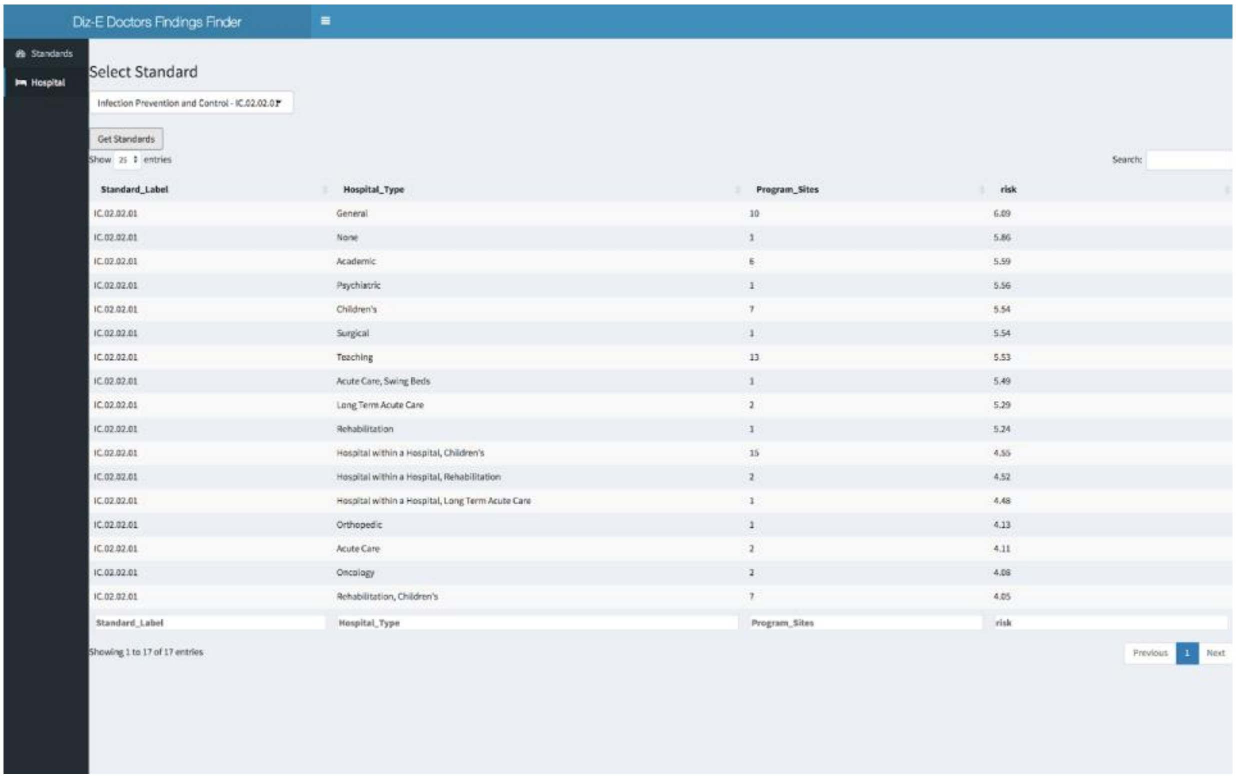


Figure **22: Screenshot of R Shiny Mobile Application Standards List**

The application is available at:

https://predict498-teamfour.shinyapps.io/health_care_dashboard/

The application runs on desktops, tablets and cell phones and connects to the Amazon RDS instance for data retrieval. Currently, the probabilities and predicted SAFER scores are saved to the database and from previous model runs. In the future, the predictions will be real-time and based on the user input. Since the R Shiny application host limits the resources and capabilities that are available for applications, the models will be run and saved on local machine then serialized. The serialized objects will then be uploaded to the R Shiny server and reloaded. This will allow the model to be used to directly make predictions without have to run it on the R Shiny server. Preliminary tests of this procedure have been positive.

CONCLUSIONS

Recommendation

We were able to generate models that used existing data to nearly perfectly predict hospital type. However, we were not able to generate a model that allowed only the findings or pattern of findings to predict hospital type in a clear, easy to use format.

The dashboard deals with the concerns in clustering and modeling by allowing one to see the probability of having one finding based on the hospital type and another finding. This will allow The Joint Commission to extend beyond the "Top 10 List" approach and allow its customers to understand the likelihood of findings across chapters.

To drill down further to the standard, a R Shiny application was developed that allows one to select the hospital type and standard finding to see the probabilities of the next 5 most likely findings and the likelihood they are found. Although a 10 percent chance of having a finding does not seem high, the random probability is 1 in 92,000.

The R Shiny application allows hospitals to prepare for surveys by addressing these findings in a controlled, planned manner rather than the fire drill post survey methodology to achieve compliance in 30 days.

The Joint Commission is also using a variation of this analysis to determine surveyor variability regarding frequency and severity of citations and has found 10 surveyors who are well outside the group. Further analysis is pending on giving them feedback.

Several data elements have been identified as required to facilitate further analysis. The Secondary Sites table provides the site ID for the program sites, but no information exists to determine the type of facility. Knowing which services the program sites provide and their size in bed count, average daily census, and visits would be useful in better predicting standards and elements of performance findings, as well as SAFER scores. Geographic information for each site would also be beneficial so the impact of location could also be evaluated. Additionally, with an identifier for each surveyor, analysis could be performed to determine if particular surveyors are associated with high SAFER scores or particular standard and elements of performance findings.

Next steps include developing models that assess whether a group of survey findings can predict a particular survey finding (e.g. do findings in Life Safety, Infection Control and Medication Management make it more likely for a finding in Leadership) and using surveyor and demographic data to assess the likelihood of a particular survey finding.

Contact Information

ANDREW BLAND MEDICAL DIRECTOR, THE JOINT COMMISSION	ERIC GERO DATA ENGINEER, GE DIGITAL	RICHARD MEYER DATA ENGINEER, SPECTRUM HEALTH	ADAM SANDSTROM DATA ANALYST, GUNDERSEN HEALTH SYSTEM	TRACY VALENTINE ACTUARY, BIG-TIME INSURANCE CO.
				
Tel 121.275.9756 abland@diz-e.com	**Tel** 121.275.9755 egero@diz-e.com	**Tel** 121.275.9752 rmeyer@diz-e.com	**Tel** 121.275.9759 asandstrom@diz-e.com	**Tel** 121.275.9750 tvalentine@diz-e.com

Company Information

Diz-E Doctors Consulting, LLC
251 E HURON ST.
CHICAGO, IL 60611
Tel 800-555-DATA
Fax 121.275.9700
www.diz-e-docs.com

Appendix A - Data File Samples

Highlighted field names are keys values used to relate files to each other.

Survey Findings			
Field Name	**SampleRow1**	**SampleRow2**	**SampleRow3**
Survey Event ID	40756097	40756097	40756097
Survey Event Program ID	40756096	40756096	40756096
Main Site ID	215866	215866	215866
Program Code	HAP	HAP	HAP
Program Type	Hospital Accreditation Program	Hospital Accreditation Program	Hospital Accreditation Program
Survey Begin Date	1/23/2017	1/23/2017	1/23/2017
Survey End Date	1/24/2017	1/24/2017	1/24/2017
Survey Year	2017	2017	2017
Standard Chapter Code	PC	PC	PC
Standard Chapter Name	Provision of Care, Treatment, and Serv	Provision of Care, Treatment, and Serv	Provision of Care, Treatment, and Serv
Standard ID	13462	13463	13639
Sstandard Label			
EP ID	55465	55472	56756
EP Label	1	1	7
Standard and EP Label	PC.03.01.07 EP 1	PC.03.01.05 EP 1	PC.04.01.05 EP 7
SAFER Score	5	5	2
SAFER Scope	Pattern	Pattern	Pattern
SAFER Likelihood	Moderate	Moderate	Low
Site ID	215866	215866	215866

Survey Event			
Field Name	**SampleRow1**	**SampleRow2**	**SampleRow3**
Main Site ID	532680	18633	210471
Program	Hospital Accreditation Program	Hospital Accreditation Program	Hospital Accreditation Program
Program Code	HAP	HAP	HAP
Hospital Type	General	Surgical	Surgical
Survey Event Type	Unannounced Full Event	Unannounced Full Event	Unannounced Full Event
Survey Begin Date	12/13/2016	6/14/2016	5/11/2015
Survey End Date	12/16/2016	6/16/2016	5/12/2015
Survey_Decision			
System Affiliation	1	0	1
Count of Non-Compliant Standards	14	23	6
Count of Non Compliant EPs	16	28	6
Total Surveyor Days	15	11	4
Number of Surveyors	5	4	2
Number of Surveyor Class Types	5	5	2
Survey Event ID	40836859	40147917	39492010
Survey Event Program ID	40836858	40147916	39492009

APPENDIX

HAP Site Detail - Main Site Characterisitics			
Field Name	**SampleRow1**	**SampleRow2**	**SampleRow3**
Main Site ID	532680	532680	532680
Hospital Type (Main Site)	General	General	General
Number of Program Sites	7	7	7
Number of DSC Certifications	0	0	0
Number of Non-DSC Certifications	0	0	0
System Affiliation	Y	Y	Y
Volume Program	HAP	HAP	HAP
Accreditation Decision	Accreditation with Full Standards Com	Accreditation with Full Standards Com	Accreditation with Full Standards Com
volume year	2017	2017	2017
update in eapp	N	N	N
Org_Volume_Category	Volumes, Hospital, Inpatient	Volumes, Inpatient, BHS	Volumes, Hospital, Outpatient
Setting	General Medical/Surgical Hospital	Inpatient	Outpatient
Volume_Name	General	Mental Health	)utpatient, Hospital, Emergency Room
Average Daily Census	83	15	0
Visits	30295	5475	40124
Beds	90	20	0

HAP Site Detail - Secondary Sites			
Field Name	**SampleRow1**	**SampleRow2**	**SampleRow3**
Main Site ID	532680	532680	532680
Site ID	334422	388316	528347
Main Site Flag	N	N	N
Miles from Main Site	77	12.2	25

APPENDIX

Appendix B - Citation

Barnett ML, Olenski AR, Jena AB. Patient Mortality During Unannounced Accreditation Surveys at US Hospitals. *JAMA Intern Med.* 2017;177(5):693–700. doi:10.1001/jamainternmed.2016.9685

Citations for Dashboard Content:

1. The Joint Commission. Facts about The SAFER™ Matrix scoring process. 4/20/2017
2. Chauhan, Susan. Course: Information Sharing: Tableau Video Tutorial. 1/12/2018

Modeling Section Citations:

Brownlee, J. (2016, September 9). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning - Machine Learning Mastery. Retrieved March 4, 2018, from https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/