Paper 1809-2014

CMS Core Measures, the Affordable Care Act, and SAS® Visual Analytics

Joe Whitehurst, High Impact Technologies; Diane Hatcher, SAS Institute

Contents

ABSTRACT

The Affordable Care Act (ACA) contains provisions that have stimulated interest in analytics among healthcare providers, especially those provisions that address quality of outcomes. High Impact Technologies (HIT) has been addressing these issues since before passage of the ACA and has a Health Care Data Model recognized by Gartner and implemented at several healthcare providers. Recently, HIT acquired SAS®Visual Analytics, and this paper reports our successful efforts to use SAS Visual Analytics for visually exploring "Big Data" for healthcare providers. Healthcare providers can suffer significant financial penalties for readmission rates above a certain threshold and other penalties related to quality of care. We have been able to use SAS Visual Analytics, coupled with our experience gained from implementing the HIT Healthcare Data Model at a number of Healthcare providers, to identify clinical measures that are significant predictors for readmission. As a result, we can help healthcare providers reduce the rate of 30-day readmissions.

INTRODUCTION

The purpose of this paper is to trace the origins of a selected subset of CMS "Core Measure", their impact on the implementation of the Affordable Care Act and show how SAS Visual Analytics can help providers track their performance on these "Core Measures" and identify important variables that impact performance so such performance can be improved thereby improving Quality of Care. The selected subset of "core measures" includes

1. Acute Myocardial Infarction
2. Heart Failure
3. Community Acquired Pneumonia

ORIGINS OF "CORE MEASURES" BY THE CENTERS FOR MEDICARE & MEDICAID (CMS) AND THEIR EVENTUAL INCORPORATION INTO THE ACA

In the spirit of sharing, which is always what these international SAS conferences have been about since the beginning for me 33 years ago when I attended my first one:
http://www.sascommunity.org/wiki/SAS_Global_Forum_Proceedings/SUGI_81

I'm going to share how we are going about developing models that can predict 30-day readmissions for specific patients by estimating their probability of being readmitted within 30 days of discharge using any and all available data from whatever sources. I'm not going to share any information about our existing models or methodologies beyond what everyone here already knows or can easily discover. In a word, these Intellectual Properties are not ready to be shared because too much work remains to be done. When they are ready to be shared, HIT will share them by publishing them. In another word, they would be too close to the farm as in...my boss says it's alright to share with colleagues, but that doesn't mean giving them the farm, just directions on how to get to the farm so they can buy or rent from the farm or its products.

Because of the enormous expense associated with early readmissions first well documented in the 1984 paper, **HOSPITAL READMISSIONS IN THE MEDICARE POPULATION,** by Anderson GF, Steinberg EP published in the New England Journal of Medicine (read abstract here), much attention has focused on early readmissions. The Anderson Steinberg paper was published 30 years ago and stimulated a flurry of initiatives to address the obvious opportunity to cut costs and improve quality of healthcare. Based on my review of the abundant literature generated by these initiatives, I think the ORXY initiative in 1987 is the most important for today's discussions because its efforts contributed so significantly to what has been incorporated into the Affordable Care Act as regards measurement methodology and reimbursement policy for hospital readmissions. It was not a shabby effort:

> ***The Hospital Core Measure Pilot Project***
>
> *Once the initial specifications for the first sets of core measure were developed, the Joint Commission initiated a pilot project to test the feasibility, usefulness, and costs associated with the implementation of core measures. Participants were drawn from eleven state hospital associations that expressed an interest in participating in a pilot project. Five were randomly selected to participate (Connecticut, Michigan, Missouri, Georgia and Rhode Island). Each hospital association then identified a single performance measurement system and 7 - 28 participant hospitals. The resulting project was a collaborative effort among the Joint Commission, five-state hospitals associations, 5 listed measurement systems, and 83 hospitals in 9 states (Connecticut, Michigan, Missouri, Georgia, Rhode Island, Texas, Virginia, California, South Carolina).*

The Joint Commission solicited ongoing feedback from participating systems and hospitals to modify core measure specifications throughout the project. Joint Commission staff also visited a random sample of 16 participating hospitals to assess the reliability of core measure data elements.

14 years after the start of the ORXY initiative and 13 years ago, they were able to report on revisions to the Core Measure Set based on what they learned from participants in the pilot study mentioned above.

History of Core Measure Set Development and Revisions

Throughout 2001, the Joint Commission received feedback from the participating state hospital associations, measurement systems, and hospitals. This feedback, the analysis of on-site reliability data, and the Joint Commission's ongoing discussions with CMS, directed toward alignment of similar measures, led to an number of improvements and modifications to the original core measures, prior to their release on November 21, 2001.

Major Revisions:

- *• Initially surgical procedures and complications were identified by key Joint Commission stakeholders as one of the initial priority areas for hospital core measure development. However, CMS is currently developing quality indicators related to surgical infection prevention (SIP) including selection and timing of prophylactic antibiotics. The Joint Commission is a member of the SIP Project Panel. Therefore, the Joint Commission will delay implementing the surgical core measure set to allow the opportunity to continue to work with CMS.*
- *• Discussions with CMS centered on measures shared between CMS and the Joint Commission and led to the development of common data element definitions and allowable values, and common measure population inclusions and exclusions.*
- *• During the pilot, data were collected for 2 measures, both of which addressed the use of angiotensin converting enzyme inhibitors (ACEI), in order to determine which of these similar measures should be implemented nationally. One measure removed the contraindications from the numerator while the other measure removed the contraindications from the denominator. Due to the complexity in the design we needed to remove the contraindications from the numerator, the cardiovascular advisory panel recommended deleting this measure in favor of the measure that was implemented nationally.*
- *• The measure pertaining to warfarin prescribed at discharge was deferred for national implementation. The measure was deferred because contraindications to warfarin are difficult and time consuming to abstract and the measure population for this indicator includes some patients not eligible for any other measures in the HF core set making it difficult to identify patients eligible for the measure.*
- *• The AMI, HF and CAP measure set all contain a measure focused on adult smoking cessation advice/counseling. The measures specifications are identical with the exception of the measure population (denominator).*

The decades of efforts by large numbers of stakeholders just outlined has brought us to where we stand today with respect to CMS “Core Measures”:

GENERAL ORYX PERFORMANCE MEASURE REPORTING REQUIREMENTS FOR MEDICAL/SURGICAL HOSPITALS

Effective with January 1, 2014 discharges, accredited general medical/surgical hospitals are required to collect and transmit data to The Joint Commission on a minimum of six core measure sets or a combination of applicable core measure sets and non-core measures as described in the table below.

Applicable core measure sets	Core measure sets required	Non-core measures required
6 core measure sets	6 core measure sets	None (date not accepted)
5 core measure sets	5 core measure sets	None (data not accepted)
4 core measure sets	4 core measure sets	None (data not accepted)
3 core measure sets	3 core measure sets	3 non-core measures
2 core measure sets	2 core measure sets	6 non-core measures
1 core measure set	1 core measure set	9 non-core measures
No core measure sets	No core measure sets	9 non-core measures

Mandatory measures sets: Four of the six measure sets are mandatory for all general medical/surgical hospitals that serve specific patient populations addressed by the measure sets and related measures. The mandatory measure sets include:

Acute Myocardial Infarction (AMI)   Pneumonia (PN)

Heart Failure (HF)   Surgical Care Improvement Project (SCIP) Perinatal Care (PC) (mandatory for hospitals with 1,100 or more births per year)

Discretionary measure sets: The sixth measure set (or fifth and sixth measure sets for hospitals with fewer than 1,100 births per year) can be chosen from among the remaining complement of core measure sets. These sets include:

Children's Asthma Care (CAC)	Stroke (STK)
Emergency Department (ED)	Substance Use (SUB)
Hospital-Based Inpatient Psychiatric Services (HBIPS)	Tobacco Treatment (TOB)
Hospital Outpatient (OP)	Venous Thromboembolism (VTE) Immunization (IMM)

As you will see in the next section, the HIT Health Data Model captures data for calculating all existing CMS "Core Measures". What I will describe now is the specific, official methodology for measuring 30-day readmissions because this clearly shows the intersection of the CMS "Core Measures" and the ACA:

> *Section 3025 of the 2010 Affordable Care Act (Public Law 111-148) requires the Secretary of the Department of Health and Human Services to establish a Hospital Readmissions Reduction Program whereby the Secretary would reduce Inpatient Prospective Payment System (IPPS) payments to hospitals for excess readmissions beginning on or after October 1, 2012 (Fiscal Year [FY] 2013).*
>
> *The Affordable Care Act further requires the Secretary to adopt the three National Quality Forum (NQF)-endorsed 30-day Risk-Standardized Readmission measures for acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN) for the Hospital Readmissions Reduction Program beginning October 2012.*
>
> *To comply with these requirements, the Centers for Medicare & Medicaid Services (CMS) will calculate Excess Readmission Ratios for these three readmission measures based on the NQF-endorsed methodology, using discharges from a prior period.*
>
> *The Excess Readmission Ratios will be used to determine the payment adjustment factors for each eligible hospital. As proposed in the FY 2014 Inpatient Prospective Payment System (IPPS) Proposed Rule, CMS intends to report these ratios in the FY 2014 IPPS Final Rule in August of 2013 as well as on the Hospital Compare website in October 2013.*

The readmission penalty in the ACA is based on readmissions for three conditions: Acute Myocardial Infarction (AMI), Heart Failure, and Community Acquired Pneumonia. For each hospital, the Centers for Medicare and Medicaid Services (CMS) calculates the risk-adjusted actual and expected readmission rates for each of these conditions. Risk-adjustment variables include demographic, disease-specific, and comorbidity factors. The excess readmission ratio is the actual rate divided by the expected rate.

Simplifying a little, the aggregate payments for excess readmissions is summed for all three conditions over the past three years, then divided by total base operating DRG payments for the past three years to calculate the penalty percentage. Base operating DRG payments are IPPS payment less DSH, IME and outliers except for new technology. The total penalty is the penalty percentage times total base operating DRG payments for that fiscal year, provided that the total amount does not exceed 1 percent of base operating DRG payments in fiscal year 2013, a cap that increases to 3 percent in fiscal year 2015 (Sahni, Cutler & Kocher, 2014).

More formerly, CMS hired **Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation** in 2008 to develop the methodologies for calculating **Excess Readmission Ratios** that will determine reimbursement penalties*(link to report)*. From the report:

> *We present a hierarchical logistic regression model for 30-day readmission after AMI hospitalization that is based on administrative data and is suitable for public reporting. The model is a strong surrogate for a similar model based on medical record data. The approach employs a grouper of 15,000+ ICD-9-CM codes that is in the public domain and yields clinically coherent variables. There is a standardized period of follow-*

up. The model does not adjust for variables that may represent complications, rather than comorbidities. The study sample is appropriately defined. The statistical approach takes into account the clustering of patients within hospitals and differences in sample size across hospitals.

The patient-level discrimination and the explained variation of the model are consistent with those observed in the development of the HF 30-day all-cause readmission measure, which was recently approved by NQF (Krumholz et al., 2008). The model performs as expected given that the risk of readmission is likely much more dependent on the quality of care and system characteristics than on patient severity and comorbidity characteristics. The readiness for discharge, the proper medications, and the proper transition to the outpatient setting may be even more important for readmission than for mortality. Results of intervention studies underscore this potential (Carroll et al., 2007; Young et al., 2003; Bondestam et al., 1995; Ades et al., 1992).

Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure. Adjusting for patient characteristics improved model performance. The ROC of 0.63 is higher than that of a model with just age and gender, 0.54, and the same as a model with all candidate variables, with ROC of 0.63. We excluded covariates, however, that we would not want to adjust for in a quality measure, such as potential complications, certain patient demographics (e.g., race, socioeconomic status), and patients' admission path and discharge disposition (e.g., admitted from, or discharged to, a skilled nursing facility). These characteristics may be associated with readmission and thus could increase the model performance to predict patient readmissions. However, these variables may be related to quality or supply factors that should not be included in an adjustment that seeks to control for patient clinical characteristics while illuminating important quality differences. For example, if hospitals with a higher share of a certain ethnic group have higher readmission rates, then including ethnic group in the model will attenuate this difference and obscure differences that are important to identify.

In summary, we present a claims-based AMI 30-day readmission measure that is suitable for public reporting. It is consistent with the consensus standards for publicly reported outcomes measures, and can be implemented using available data.

From that beginning, the **Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation** provides an annual update. An excerpt from the latest update will explain why one should also consult the latest update before working with these selected CMS "CORE Measures, **AMI, HF, and pneumonia readmission measures**":

This report describes three of the Centers for Medicare & Medicaid Services (CMS) readmission measures used in the Hospital Inpatient Quality Reporting (IQR) program and publicly reported on Hospital Compare: the hospital-level 30-day risk-standardized readmission rates (RSRRs) following acute myocardial infarction (AMI), heart failure (HF), and pneumonia measures.
This report is intended to provide a single source of information about the current measures for a wide range of readers. Within this report we provide an overview of the measure methodology, describe methodology updates to the measures and the national results for 2013 public reporting, and describe our quality assurance processes. The appendices provide further details, including concise tables of measure specifications and a list of the annual updates each year since public reporting began in 2009.

Specifically, the reader can find:

An overview of the AMI, HF, and pneumonia readmission measures (Section 2):

History of the measures

Measure cohort included and excluded hospitalizations

How transfers are handled

differences between IQR reporting and the Hospital Readmissions Reduction Program (Section 3025 of the Affordable Care Act)

Outcome (what counts as a readmission) what is considered a planned readmission

Risk-adjustment specifications

Data sources

Readmission rate calculation

Categorization of hospitals' performance

2013 measure updates (Section 3): *The most significant update for 2013 reporting is the addition of an algorithm to identify planned readmissions. Planned readmissions will not be counted in the measures.*

2013 results (Section 4)*: Results from the models that are used for the Hospital Inpatient Quality Reporting (IQR) program in 2013.*

Quality assurance process (Section 5)

The Appendices contain detailed measure information, including:

Appendix A: Measure specifications;

Appendix B: Annual updates to measures since measure development;

Appendix C: Detailed overview of the Planned Readmission Algorithm;

Appendix D: Definitions for common terms; and

Appendix E: RTI's memorandum on updates to the Condition Category (CC) map.

RELEVANCE OF THE HIGH IMPACT TECHNOLOGIES HEALTHCARE DATA MODEL FOR IMPLEMENTING AND COMPLYING WITH THE ACA

HITACTICS™ HEALTH DATA MODEL HIGHLIGHTS

A solid foundation for turning health care data into strategic insights

- Turn operational data into strategic insight with end-to-end integration
- Adapt to evolving requirements to maximize reimbursements, compliance and profitability
- Expand analytical dashboards and reports to include emerging clinical areas
- Increase agility and decrease time for delivering new reports
- Align business and technical resources with a common vocabulary to accelerate new initiatives

Healthcare leaders are under tremendous pressure to reduce cost, improve outcomes for patients and provide more coordinated and personalized care. Some of these benefits will emerge as the momentum to transform healthcare organizations into more patient-centric and collaborative entities continues to build. In particular, the 2009 Patient Protection and Affordable Care Act (ACA), is helping to accelerate the pace of these demands. While offering key incentives for change, the act also provides provisions to drive down the cost of healthcare. In fact, most of the innovative health care provider organizations we work with expect their revenue to go down 10% to 20% over the next ten years.

Faced with these challenges, healthcare providers are increasing their investments in key information management and business analytics systems. They are seeking to gain critical business insights by implementing enterprise data warehouses and to integrate information across their enterprises and enable meaningful measurements and effective decision making.

Meeting the need for better access to integrated health care data

Achieving analytical value from health care data requires comprehensive data across clinical, financial and operational data elements along with the ability to leverage existing technologies and data schemas. The HITactics™ Health Data Model delivers the blueprint that health care providers need to rapidly deploy a broad range of analytics and to realize the insights that will help them adapt to the new health care industry environment.

Implementing a data warehouse quickly and cost-effectively

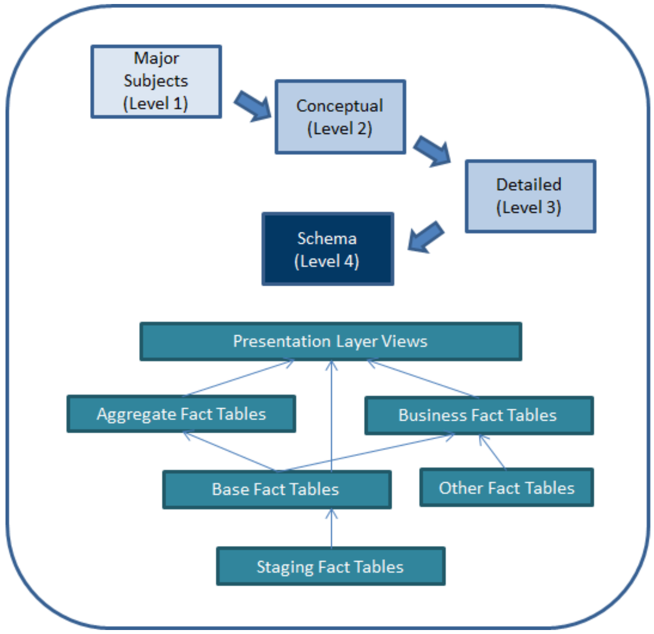
Healthcare providers can use the HITactics Health data model to correlate all of their data in a cohesive and flexible framework. It is the result of over 20 years of industry experience across hundreds of client engagements. The models modular design allows IT personnel to deploy dynamic, analytical data stores that combine both internal and external data quickly. This helps decision makers better understand health trends within specific populations and identify opportunities for innovation – all at a reasonable cost.

Key components of the Healthcare Provider Data Model include standard business terms to help ensure a common understanding between business users and IT, a library of BI templates to demonstrate typical reporting and analysis requirements and a set of data models by subject area.

Business Data Model: The BDM is a logical entity-relationship (ER) model that represents the essential entities and relationships within health care.

Dimensional Model: The DWM is a physical representation of the business data model.

Atomic Model: This is a logical representation of the business data model, optimized for data repositories that need specific, efficient structures dedicated to the storage of historical facts.



Staging Layer (Tables)

- Import tables from various source systems
- Temporary storage for each load cycle

Physical Layer (Tables)

- Permanent data structures for dimensional and fact data
- Snowflaked-star schema design
- Common location for data across source systems
- Optimized for data warehousing and analytics

Presentation Layer (Views)

- Used as source for BI metadata
- Tables and columns can be combined, duplicated and/or renamed

Business Layer (Tables)

- Common healthcare calculations of data
- Data may come from multiple source tables
- Pre-calculated in ETL
- Data stored at lowest possible level for detailed reporting
- Results are also aggregated to improve performance
- Common candidates: Quality Measures, Meaningful Use

Aggregate Layer (Tables)

- May vary depending on specific database platform
- Include aggregate tables on the largest fact tables
- Can only be done with ‘aggregatable’ fact tables
- Used to reduce rows (‘width’) and columns (‘depth’) of tables
- Include multiple versions/layers per fact table
- Common candidates: Charges, Financial Transactions

HITactics Health Modules
Foundation Module[1]
Clinical
Perioperative
Emergency Department
Pharmacy
Scheduling
Key Performance Indicators
Quality Measurement
Population Health
Finance (G/L, A/R, A/P, F/A)
Revenue Cycle
Physician Practices
Human Resource
Supply Chain

The Foundation module is comprised of the portions of the data model that are required by and common to the other modules. This includes components like standard hierarchies (e.g. organization, geography, date) and dimensions (e.g. CDS, charge, CPT, diagnosis, DRG, patient, physician, service, resource, quality measure, procedure & G/L account). It also includes the Patient Encounter or Visit tables which are required by most of the other modules.

ORIGINS OF VISUAL EXPLORATORY DATA ANALYSIS—SAS® VISUAL ANALYTICS

Many recognize John Tukey as the father of exploratory data analysis, in part for creating many effective visual techniques such as the box plot and stem and leaf plot, which are standards in introductory statistics courses today. He made many enduring contributions in time series, multiple comparisons, ANOVA, robust statistics, and interactive and multivariate graphics, too. And, he is recognized as the father of exploratory data analysis because of his seminal work on the subject, Exploratory Data Analysis published by Addison-Wesley in 1977.

John Tukey proposed a new approach to data analysis, based heavily on visualization, as an alternative to classical (mathematical) data analysis. Being dependent on graphics, this approach only became practical with the advent of modern computers. However, in addition to advocating the graphical techniques of visual data analysis, he proposed the methodology of data exploration, a methodology in which a model of the phenomena might be inferred instead of pre-imposed. It is this powerful combination that led him to coin the phrase "exploratory data analysis", commonly referred to simply as "EDA." This exploratory approach is appropriate as a first step in data mining because it allows one to explore data with an open mind. The graphical techniques of visual exploration, in combination with humans' natural pattern-recognition capabilities and knowledge of the subject, facilitate the discovery of the structural secrets of the data.

Tukey suggests that one think of EDA as the first step in a two-step process similar to that utilized in criminal investigations. In that first step, one searches for evidence using all of the investigative tools that are available. In the second step, that of confirmatory data analysis, one evaluates the strength of the evidence and judges its merits and applicability. It is in this second step that one would likely evaluate the model(s) which have been inferred during exploration and likely apply the techniques of classical data analysis.

He considered himself a "scientific generalist" and influenced many other scientific fields, including computer science, mathematics, engineering and economics. He even served as an adviser on environmental, defense and education policy to the highest levels of US government. He collaborated with scientists such as Von Neumann, Feynman, Cooley and Morganstern, and served as doctoral adviser to many great statisticians (Hinrichs, 2013).

Visualization is critical for quickly understanding what's happening in your data and seeing things you couldn't before. But while data visualization enables you to explore your data, analytic visualization helps you discover insights buried in it. Because of this, analytic visualizations provide more value than simply exploring and using data visualization techniques.

The use of visual analytics is growing in the health care industry by helping organizations quickly and efficiently sort through huge volumes of unstructured and structured data. Graphical dashboards bring data to life by presenting essential information in an automated fashion. This significantly helps decision makers view data in familiar and user-friendly formats—charts, graphs, tables, pictograms and others. According to estimates from market researcher Frost & Sullivan, even a modest-sized hospital's electronic medical records can easily run into multiple terabytes, and it wouldn't be surprising to see some larger hospitals with petabyte-level storage requirements just for patient record. Data visualization effectively reduces massive and complex data structures into a format that better aligns with the brain's ability to recognize, analyze and act on information. Current state-of-the-art visualization techniques, when added to deep analytics, make it easier for users to register data values than they could when leafing through spreadsheets or other traditional data-reporting techniques.

("Unleashing the Value of Data Hidden Within Health Care Organizations", Feb 2014,

http://www.healthdatamanagement.com/digital_edition/unleashing-the-value-of-data-46922-1.html

SAS VISUAL ANALYTICS

Published just one year after SAS was incorporated in 1976, Tukey's Exploratory Data Analysis undoubtedly had an early and significant influence on the development of SAS software. Commonly used descriptive statistics and exploratory printer graphics suitable for analyzing continuous variables are among the first procedures developed by SAS. PROCs MEANS, UNIVARIATE and PLOT (now GPLOT) provide a wide range of summary and exploratory statistics. Tukey's work had a profound influence on the development of SAS Visual Analytics

SAS Visual Analytics provides a complete platform for analytics visualization, combining an easy-to-use, dynamic interface with powerful in-memory technology. This enables you to identify patterns and relationships in data that weren't evident before.

- Interactively explore data
- Execute analytic correlations, forecasts or decision trees on any size data within seconds
- Deliver results quickly wherever needed – either on the web or to mobile devices

Interactive, self-service BI and reporting capabilities are combined with out-of-the-box advanced analytics to help you discover insights from any size and type of data, including text.

After describing some of Tukey's most important contributions, we will address how we have incorporated SAS Visual Analytics—what's been done and what is planned for the future. We will start with Tukey's 5-number summary statistics.

FIVE-NUMBER SUMMARY STATISTICS

The five number summary of a continuous variable consists of the minimum value, the first quartile, the median, the third quartile, and the maximum value. The median, or second quartile, is the mid-value of the sorted data. The first quartile is the 25th percentile and the third quartile is the 75th percentile of the sorted data. The range between the first and third quartiles includes half of the data. The difference between the third quartile and the first quartile is called the inter-quartile range (IQR). Thus, these five numbers display the full range of variation (from minimum to maximum), the common range of variation (from first to third quartile), and a typical value (the median).

Measures of Dispersion

1. **Range.** Range is the difference between the maximum and minimum values. It is easy to compute because only two values, the minimum and maximum, are used in the estimation; however, a great deal of information is ignored, and the range is greatly influenced by outliers.
2. **Variance.** Variance is the average measure of the variation. It is computed as the average of the square of the deviations from the average; however, because variance relies on the squared differences of a continuous variable from the mean, a single outlier has a greater impact on the size of the variance than does a value close to the mean.
3. **Standard Deviation.** The Standard Deviation is the square root of the Variance. In a normal distribution, about 68% of the values fall within one standard deviation of the mean, and about 95% of the values fall within two standard deviations of the mean. Both variance and standard deviation measurements take into account the

difference between each value and the mean. Consequently, these measures are based on a maximum amount of information.

4. **Inter-quartile range.** The IQR is a robust measure of dispersion. It is the difference between the 75th percentile (Q3) and the 25th percentile (Q1). The IQR is hardly affected by extreme scores; therefore, it is a good measure of spread for skewed distributions. In normally distributed data, the IQR is approximately equal to 1.35 times the standard deviation.

Standard Errors and Confidence Interval Estimates

1. Standard error. Standard error is the standard deviation of the sampling distribution of a given statistic. Standard errors show the amount of sampling fluctuation that exists in the estimated statistics in repeated sampling. Confidence interval estimation and statistical significance testing are dependent on the magnitude of the standard errors. The standard error of a statistic depends on the sample size. In general, the larger the sample size, the smaller the standard error.

2. Confidence Interval. The confidence interval is an interval estimate that quantifies the uncertainty caused by sampling error. It provides a range of values, which are likely to include an unknown population parameter, as the estimated range is being calculated from a given set of sample data. If independent samples are taken repeatedly from the same population, and a confidence interval is calculated for each sample, then a certain percentage of the intervals will include the unknown population parameter. The width of the confidence interval provides some idea about the uncertainty of the unknown parameter estimates. A very wide interval may indicate that more data must be collected before making inferences about the parameter.

Detecting Deviations from Normally Distributed Data

1. **Skewness.** Skewness is a measure that quantifies the degree of asymmetry of a distribution. A distribution of a continuous variable is symmetric if it looks the same on the left and right of the center point. Data from positively skewed (skewed to the right) distributions have values that are clustered together below the mean and a long tail above the mean. Data from negatively (skewed to the left) distributions have values that are clustered together above the mean but have a long tail below the mean. The skewness estimate for a normal distribution equals zero. A negative skewness estimate indicates that the data are skewed left (the left tail is heavier than the right tail), and a positive skewness estimate indicates that the data are skewed right (the right tail is heavier than the left tail).

2. **Kurtosis.** Kurtosis is a measure to quantify whether the data are peaked or flat relative to a normal distribution. Datasets with large kurtosis have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Datasets with low kurtosis have a flat top near the mean rather than a sharp peak. Kurtosis can be both positive and negative. Distributions with positive kurtosis have typically heavy tails. Kurtosis and skewness estimates are very sensitive to the presence of outliers. These estimates may be influenced by a few extreme observations in the tails of the distribution; therefore, these statistics are not a robust measure of non-normality. The Shapiro-Wilks test and the d'Agostino-Pearson omnibus test are commonly used for detecting nonnormal distributions.

Graphical Techniques Used in EDA of Continuous Data

Graphical techniques convert complex and messy information in large databases into meaningful displays; no quantitative analogs can give the same insight as well-chosen graphics in data exploration.

1. **Frequency histogram.** The horizontal frequency histogram displays classes on the vertical axis and frequencies of the classes on the horizontal axis. The frequency of each class is represented by a horizontal bar that has a height equal to the frequency of that class.

2. **Box plot.** A box plot provides an excellent visual summary of many important aspects of a distribution. The box plot is based on the five-number summary plot, which is based on the median, quartiles, and extreme values. The box stretches from the lower hinge (first quartile) to the upper hinge (the third quartile) and therefore contains the middle half of the scores in the distribution. The median is shown as a line across the box. Therefore, one quarter of the distribution is between this line and the top of the box, and one quarter of the distribution is between this line and the bottom of the box. A box plot may be useful in detecting skewness to the right or to the left.

3. **Normal probability plot.** The normal probability plot is a graphical technique for assessing whether or not a variable is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality. A normal probability plot, also known as a normal Q-Q plot (or normal quantile-quantile plot), is the plot of the ordered data values (y axis) against the associated quantiles of the normal

distribution (*x* axis). For data from a normal distribution, the points of the plot should lie close to a straight line. Normal probability plots may also be useful in detecting skewness to the right or left.

4. **2D and 3D Scatterplots.** Scatter plots are useful for displaying the relationship of 2 (3 in 3D plots) variables and are indispensable tools in model building. At a glance, skewness, departures from normality and relationship between 2 variables can be seen

Data Exploration: Categorical Variables

One-way and multi-way frequency tables of categorical data are useful in summarizing group distributions and relationships between groups and for checking for rare events.

Descriptive Statistical Estimates

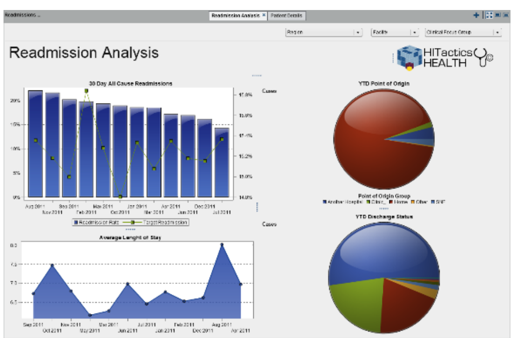
1. **Cross tabulation.** Cross tabulation uses a two-way table to show the frequencies for each level in one categorical variable across the levels of other categorical variables. One of the categorical variables is associated with the columns of the contingency table, and the other categorical variable is associated with the rows of the contingency table. This table is commonly used to display the association between two categorical variables.

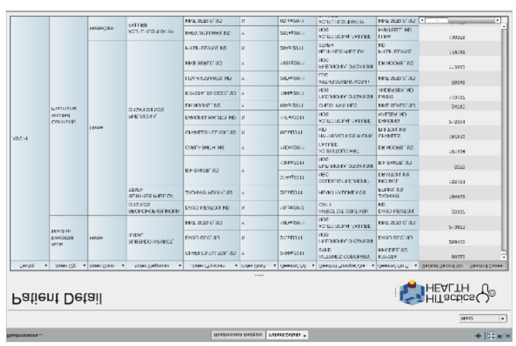
2. **Pearson's chi-square test for independence.** For a contingency table, Pearson's chi-square test for independence tests the null hypothesis that the row classification factor and the column classification factor are independent by comparing observed and expected frequencies. The expected frequencies are calculated by assuming that the null hypothesis is true. The chi-square test statistic is the sum of the squares of the differences between the observed and expected frequencies, with each squared difference being divided by the corresponding expected frequency.

READMISSIONS ANALYSIS DASHBOARD

Dashboard Overview:

- Summarizes 30-day all cause readmission information for the following three key clinical focus groups (CFG):
  - Acute Myocardial Infarction
  - Heart Failure
  - Community Acquired Pneumonia
- Allows business users to “slice & dice” information at the organization, region or facility level
- Provides ability to drill into patient details for context on readmissions





Key Features / Components:

- *30-Day All Cause Readmissions* - trends and compares the last 12 months of actual readmission rates (blue bars) to the organization’s target readmission rate (red line)
- *Average Length of Stay* (ALOS) – 12 month trend of patient ALOS (for the index visit prior to readmission)
- *YTD Point of Origin* – year-to-date readmission cases broken down by point of origin (i.e. where patient was readmitted from)
- *YTD Discharge Status* – year-to-date readmission cases broken down by discharge location (i.e. where patient was discharged to)

- *Drill to Patient Details* – click (web) or tap (iPad) on any bar of the 30-day All Cause Readmissions graph to drill into patient details via a pop-up information window

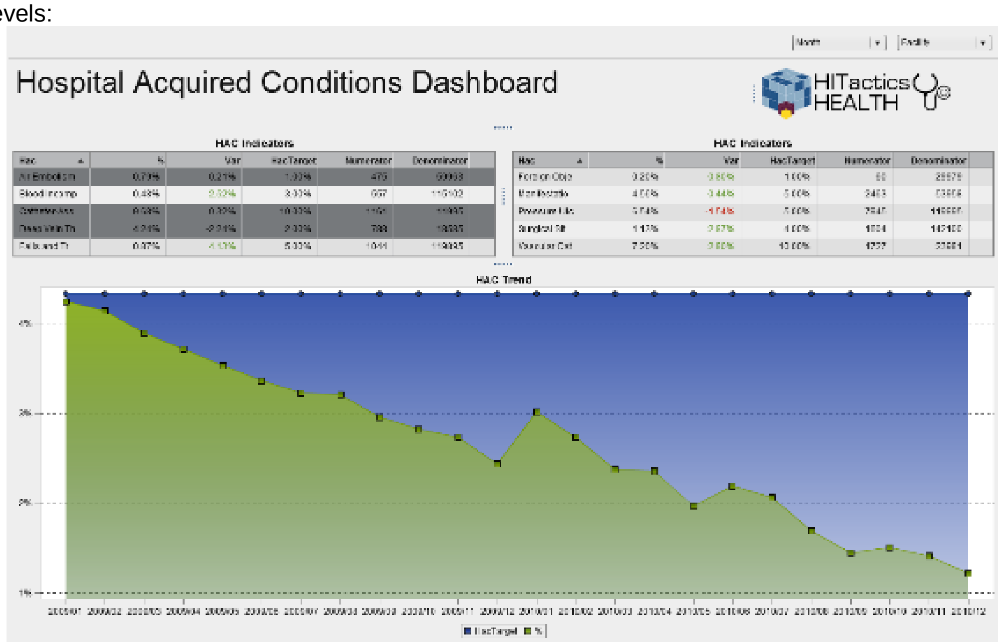
Additional Supplementary Reports Provided:

- 30 Day All Cause Readmissions
- 30 Day All Cause Readmission Rates by CFG
- Readmission ALOS
- Readmission Point of Origin
- Readmission Discharge Status
- Readmission Patient Details

HOSPITAL ACQUIRED CONDITIONS DASHBOARD

Dashboard Overview:

Graphical comparative analysis and trending of hospital acquired conditions (HAC) rates at enterprise, regional and facility levels:



Key Features / Components:

- *HAC Indicators* – Users are able to reviewing and compare the following statistics associated with the hospital acquired conditions indicators:
    - *%* – Percentage / rate of a specific hospital acquired condition
    - *Variance* – Green or red arrow icon which denotes a positive or negative variance from the HAC target
    - *Numerator* – Total cases with a specific hospital acquired condition
    - *Denominator* – Total population at risk for a hospital acquired condition
- *HAC Trend* - Clicking (web) or tapping (iPad) on any one of the hospital acquired conditions in the detail grid updates the % vs. target trend graph at the bottom of the screen
- *Date / Time Slider* - Allows users to interact / manipulate the time series associated with the HAC Trend graph

Additional Supplementary Reports Provided:

- HAC Measures
- HAC Measures Variance
- HAC Measures Variance Trend

QUALITY MEASURES DASHBOARD

Dashboard Overview:

Graphical comparative analysis and trending at enterprise, regional and facility levels of compliance with the following five clinical quality measure classes:

- Acute Myocardial Infarction (AMI)
- AMI & Chest Pain
- Heart Failure
- Pneumonia
- Surgical Care Improvement Program (SCIP)

Key Features / Components:

- *Quality Measures Statistics* – Users have the option of reviewing the statistics associated with all Quality Measures or only those specific to a single indicator class
- *Quality Measure Trends* - Clicking (web) or tapping (iPad) on any one of the Quality Measures in the detail grid updates the ratio vs. target trend graph at the bottom of the screen
- *Date / Time Slider* - Allows users to interact / manipulate the time series associated with the Quality Measures Trend graph

Additional Supplementary Tabular/Grid Reports Provided:

- Quality Measures

MEANINGFUL USE DASHBOARD

Dashboard Overview:

Graphical comparative analysis and trending of compliance with Meaningful Use (MU) core and menu objectives at enterprise, regional and facility levels

Key Features / Components:

- *MU by Measure Panel* – Users have the option of reviewing the statistics associated with all or one specific MU Measure for each facility within the selected region.
- *MU by Facility Panel* - Users have the option of reviewing the statistics associated with all or one specific facility for each MU Measure within the selected measure type.
- *Graphical Comparisons & Trends* – Clicking (web) or tapping (iPad) on any of the Facilities (By Measure Panel) or Measures (By Facility Panel) detail grid updates the graphical segment at the bottom of each panel. In addition, users have the option to view the MU information graphically by:
    - *Current Ratio vs*. Target – Compares current MU measure rate to the MU target (by facility or measure)
    - *YTD Ratio vs*. Target – Compares the YTD MU measure rate to the MU target (by facility or measure)
    - *Trending* – Line graph trending of the MU measure rate (by facility or measure) with ability to adjust trend time period with date slider bar

Additional Supplementary Reports Provided:

- Meaningful Use – Key Metrics
- Meaningful Use – Key Metrics by Nurse Station
- Meaningful Use – Key Metrics by Attending Physician

CONCLUSION

It is too early in this process to have any conclusions. What we have tried to do is set the stage for understanding why we want to use SAS Visual Analytics to explore voluminous healthcare data. During the presentation, we will present many examples of visualizing healthcare data that may have relevance for developing models to predict 30-day readmissions for patients admitted for Acute Myocardial Infarction (AMI), Heart Failure, or Community-Acquired Pneumonia. We will also show you an expanded, detailed view of the HIT Health Data Model for a few minutes.

REFERENCES

A Comprehensive Review of Development and Testing for National Implementation of Hospital Core Measures Available athttp://www.jointcommission.org/NR/rdonlyres/48DFC95A-9C05-4A44-AB05-1769D5253014/0/AComprehensiveReviewofDevelopmentforCoreMeasures.pdf

Anderson GF, Steinberg EP Hospital readmissions in the Medicare population
The New England Journal of Medicine [1984, 311(21):1349-1353]

Chris Feudtner, MD, PhD, MPH, James E. Levin, MD, PhD, Rajendu Srivastava, MD, MPH, Denise M. Goodman, MD, MS, Anthony D. Slonim, MD, DrPH, Vidya Sharma, MBBS, MPH, Samir S. Shah, MD, MSCE, Susmita Pati, MD, MPH, Crayton Fargason Jr, MD, MBA, and Matt Hall, PhD How well can hospital readmission be predicted in a cohort of hospitalized children? A retrospective multi-center study
Pediatrics. 2009 January ; 123(1): 286–293. doi:10.1542/peds.2007-3395.

Desai, M. M., Lin, Z., Schreiner, G. C., et al. 2009 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures: Report prepared for the Centers for Medicare & Medicaid Services. 2009; Available at: http://www.qualitynet.org/

Issac Shams ,2012, USING THE SAS® SYSTEM TO DEVELOP RISK PREDICTION MODELS FOR PATIENT RE-ADMISSION REDUCTION IN VAMCS http://www.misug.org/uploads/8/1/9/1/8191072/ishams_readmission.pdf

Jian Dai, Zhongmin Li, David Rocke
University of California, Davis, CA Hierarchical Logistic Regression Modeling with SAS GLIMMIX
http://www.lexjansen.com/wuss/2006/analytics/ANL-Dai.pdf

Krumholz HM, Normand S-LT, Keenan PS, et al. Hospital 30-Day Heart Failure Readmission Measure: Methodology. Report prepared for the Centers for Medicare & Medicaid Services. 2008; Available at: http://www.qualitynet.org/

Krumholz HM, Normand S-LT, Keenan PS, et al. Hospital 30-Day Acute Myocardial Infarction Readmission Measure: Methodology. Report prepared for the Centers for Medicare & Medicaid Services. 2008; Available at: http://www.qualitynet.org/

Krumholz HM, Normand S-LT, Keenan PS, et al. Hospital 30-Day Pneumonia Readmission Measure: Methodology. Report prepared for the Centers for Medicare & Medicaid Services. 2008; Available at: http://www.qualitynet.org/

Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. 2006; 113: 456-462.

Kansagara D, Englander H, Salanitro A, et al. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA* 2011; 306(15): 1688-1698.

Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M and Kripalani S. Risk Prediction Models for Hospital Readmission: A Systematic Review. VA-ESP Project #05-225; 2011.

Liu L, Forman S, Barton B 2009. Fitting Cox Model Using PROC PHREG and Beyond in SAS. SAS Global Forum. Available from: http://support.sas.com/resources/papers/proceedings09/236-2009.pdf

Marc D. Silverstein, MD, Huanying Qin, MS, S. Quay Mercer, MT(ASCP), Jaclyn Fong, MPH, and Ziad Haydar, MD Risk factors for 30-day hospital readmission in patients ≥65 years of age, Proc (Bayl Univ Med Cent). 2008 Oct;21(4):363-72.

Powell T, Bagnell M 2012. Your “Survival” Guide to Using Time-Dependent Covariates. SAS Global Forum. Available from: http://support.sas.com/resources/papers/proceedings12/168-2012.pdf

Sahni, Cutler & Kocher 2014. Will The Readmission Rate Penalties Drive Hospital Behavior Changes? ***Health Affairs***

Tukey, J.W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Joe Whitehurst
Organization: High Impact Technologies
2002 Summit Boulevard NE, Suite 3100
Atlanta, Georgia 30319
Phone: 1 404.460.7001
Email: jwhitehurst@hitactics.com
Web: www.hitactics.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.