Validating the PRIDIT Method for Determining Hospital Quality with Outcomes Data

Sponsored by
Society of Actuaries
Health Section

Prepared by
Robert D. Lieberthal
Dominique M. Comer
March 2013

Foreword

We would like to acknowledge the people who participated in the development of this research report.

We would also like to thank the Society of Actuaries Project Oversight Group (POG), which provided advice and guidance during the course of the project. We also would like to thank the research coordinators from the Society of Actuaries for their efforts in organizing and coordinating the project.

POG members:

Cabe Chadick

Lijia Guo

Song-Song Liao

Jed Linfield

Michael Rasmussen

Sandee Schuster

Jon Shreve

Bram Spector

Sara Teppema

Junjie Zhang

SOA project coordinators:

Barbara Scott

Steven Siegel

Jefferson School of Population Health researchers:

Katie O'Connell

Other acknowledgements:

Part of Dr. Comer's time spent on this research was funded by a Postdoctoral Fellowship award on Health Outcomes from the PhRMA Foundation.

Executive Summary[1]

Hospital Quality through the Eyes of PRIDIT

What makes PRIDIT shine?

With the great amount of data that is collected from hospitals, it can be difficult to determine which data are important to collect and what the data truly says about the quality of a hospital. Much of this within a hospital is easily observable and includes the following types of data:

- Process measures (i.e. what the hospital staff actually does during the course of care);
- Structural measures (i.e. the physical or service capabilities of the hospital and its staff);
- Outcomes measures (i.e. the results and endpoints of care, such as mortality or readmission);
- Consumer satisfaction measures of patient experience in a hospital

The PRIDIT method is able to combine different types of data to create an overall picture of hospital quality. How well hospitals score on the measures they report on will determine the quality of the hospital. Measures that are individually important, good indicators of performance within a measure type, and good indicators of performance across all measure types will tend to get the highest weights. Similarly, PRIDIT accounts for structural data that will affect multiple types of performance, such as not-for-profit status. Ultimately, PRIDIT also will rank hospitals based on their performance on these measures.

PRIDIT demonstrates the value in utilizing many types of data to determine proxies for hospital quality to focus on the specific drivers of quality. At a single point in time, a small number of measures account for the largest part of the quality scores. The combination of data through PRIDIT shows the possibility for prioritizing quality measures, both at a single point in time and across time. Hospitals should focus on collecting the data measures that will be most useful; PRIDIT can point to these measures. We show that drivers are largely consistent over time. Thus, data containing the most useful measures will also tend to consist of the same measures over time.

Another use for PRIDIT is to observe the comparative results of the hospitals surveyed. The advantage of looking at the distribution of scores is that PRIDIT, when applied to our rich data set for a national set of hospitals, is less sensitive to the scores or patient counts of individual hospitals. The distribution can show how many truly great hospitals there are, as well as outliers that researchers

[1] This section includes significant input from Cabe Chaddick.

identify and study for further lessons learned. It can show whether the distribution of scores is continuous or whether there are certain specific quality "types", such as high, average, and low quality, that are distinct from each other. It can also show over time whether the distribution of hospital quality is changing and how. The statistical distribution of quality may look like a Normal distribution, a highly skewed distribution, or some other curve. These distributions would imply very different numbers of very high quality hospitals, among other implications.

Results of the Analysis

Distribution of quality

Overall, the dispersal of hospital quality is fairly even, with a slight tendency for hospitals to be worse than average. Most hospitals' scores fall into the two-standard deviation range of (−0.015, 0.015), so the distribution can be fit to standard models commonly used in actuarial science. The results show very few hospitals are of extremely high or extremely low quality. Hospitals with lower mortality rates tend to receive better quality scores compared to others. Hospitals that saw more patients also receive better scores, meaning practice makes perfect. For hospital quality, the more an individual or team repeats an action, the more proficient they become in that action (also known as the "volume-outcome relationship"). Of important note, higher consumer satisfaction scores are negatively correlated with quality. This may imply that the hospitals with qualities that patients favor (i.e., quietness) may not necessarily be of high quality. A surprising finding from this study is that hospitals with higher readmission rates tend to be of higher quality, as has been found in a small number of other studies. This adds to the literature that argues against programs that penalize hospitals for high readmission rates.

Quality variance across geography and its use

Our results show that there is a lot of variation in care in regions of the U.S. with a high concentration of providers. In these regions, improving hospital quality could involve emulating the highest quality hospital, or even individuals or insurers shifting their business such that lower quality providers consolidate or close. The availability of quality hospitals in regions with many competing providers allows insurers and consumers the opportunity to travel to higher quality providers, using the results of our study to determine which hospitals have superior quality.

In low-density areas, improving the single hospital available is the only way to ensure quality. While some individuals are fortunate enough to live near an "island of quality," others must accept the level of care they live near or travel a great distance to gain access to higher quality care. For such areas with lower levels of quality, building a new facility is likely not an option. Rather, the focus must be on improving the hospital by focusing on the most impactful measures as can be identified by PRIDIT. In the extreme, individuals or insurers shifting their business such that the lower quality hospital closes may be an option, as it would result in higher volumes at the remaining facilities. In this manner, the PRIDIT method shows several ways to improve quality, and thus provides options for quality improvement efforts.

Our analysis identified certain geographical areas with many higher quality hospitals. The top 100 hospitals identified from our analysis are concentrated in the New York City—Washington DC corridor and the upper Midwest, with a smaller number in Florida and California. Contrary to our original hypothesis, we observed that teaching hospital variables are not amongst the top 20 of variable rankings. No single variable or group of variables causes this geographic clustering of higher quality. Instead, it is a preponderance of a number of variables. The higher quality hospitals in these geographic areas have higher scores in most but all of the contributing variables.

Quality improvement

Our results also suggest that hospitals seeking to learn from higher quality facilities have several options. One is to identify the few very high quality hospitals, and learn from them. Another is to find a comparator hospital of higher quality that is realistically achievable. In our analysis of PRIDIT scores, we found that for the hospitals in general, their scores are continuous. In other words, almost any hospital, except for the very best, should be able to identify a significant number of comparable but higher quality hospitals using our results. Every quality score has a near neighbor, and quality scores do not appear in clumps. Then, these higher quality comparator hospitals could offer lessons learned that would not require wholesale change. That is a result of the fact that the hospital scores are continuous, or in other words not discrete.

This result reaffirmed our decision to group all hospitals together in one consolidated group regardless of type (e.g., academic medical centers, rural community hospitals, Veterans Affairs hospitals, etc.) and then score them against one another within this aggregated group. In other words, we do not need separate methods by hospital type to score. That is for the positive since the more comparator hospitals, the better the data is, and the more potential there is for lessons learned to apply to hospital quality improvement—rural critical access hospitals could learn from a Veterans Affairs hospital, for example.

PRIDIT is stable. That is, hospital quality scores in one year are a likely indicator that that hospital will produce a similar quality score the next year. We found that PRIDIT scores predict quality for up to two years. As a result, patients, providers, and payers using PRIDIT to make choices for the future can be relatively sure that they are using a stable measure of quality.

Actuarial Role

Actuarial implementation of PRIDIT could allow actuaries to differentiate quality of hospitals within a market or between markets for hospital care. It allows for the benchmarking of hospitals, locally and nationally, on quality scores. Insurance companies now have Pay-for-Performance (P4P) programs that collect similar data from hospitals. By applying PRIDIT, insurance companies could pay for improvements in PRIDIT scores rather than individual measures, or use PRIDIT scores to pay outperforming hospitals. Hospitals might also benefit from comparing their PRIDIT scores based on data from each insurance company, and government payers, in order to target quality improvement efforts.

Market dynamics, rather than hospital quality or outcomes, may determine provider reimbursement. It could be the case that higher quality care is costlier. In that case, individuals living in an area with better hospitals would receive better care and thus be costlier to insure. The opposite could also be true. Researchers should further explore this relationship.

An additional insurance application is that of preferred provider networks. This should be a direct consequence of having different quality levels in a community large enough to support multiple providers. An insurer could select a level of quality that it wishes to associate with a particular health insurance plan. Then, it could selectively contract with the providers at that level. Such an insurer could even offer plans with different quality levels, much as they now offer plans with different deductibles or different designs. PRIDIT gives an objective way to set up, and adjust, such levels over time.

In conclusion, the PRIDIT method shows several ways to improve quality, and thus provides options for quality improvement efforts. Actuaries can use the PRIDIT method to understand and communicate hospital quality to other stakeholders including providers, employers, and policymakers. Utilizing this method can improve hospital quality as well as patient outcomes in an efficient way, allowing analysts to construct a single score of hospital quality to increase the value of health insurance and improve patient outcomes. PRIDIT also has a potential role in other health care settings, such as outpatient clinics and pharmacies. Given the findings of this study, researchers in the health care setting should also further explore the prior use of PRIDIT as a fraud detection method.

Table of contents

I. Background[2]

A. Introduction to measuring and defining hospital quality

What deems a hospital to be "high quality?" Low mortality rates? Accreditation? Teaching status? While many hospitals strive for each of these goals, to what degree do they reflect the quality of a hospital? Determining the objective quality of a hospital has proved to be a difficult and controversial topic (Lovaglio, 2012; Shahian, Wolf, Iezzoni, Kirle, & Normand, 2010). As health care reform moves forward and our focus shifts to health outcomes and more efficiency, we now place even more importance on the performance of our hospitals. Programs such as Pay for Performance and Meaningful Use collect measures from hospitals in an attempt to improve the quality of care. Organizations such as the Leapfrog Group (The Leapfrog Group, 2010) create public report cards to allow for the direct comparisons of hospitals and specialty clinics. The critical piece that is missing from all of these initiatives is that they do not quantify the degree to which different factors contribute to quality. The PRIDIT method attempts to quantify quality, based on the conglomeration and relative ranking of measures collected by Hospital Compare (through the Center for Medicare and Medicaid Services).

Hospitals provide a variety of services and perform a number of different functions therein, each of which contribute to the overall quality of that hospital. The true quality of a hospital stems from a multifactorial process, in which both measurable and non-measurable factors tie into the health outcomes and satisfaction of patients. Example factors of quality can range from the amount of time a nurse spends with a patient to mortality rates to post-operation infection rates; these factors are directly measurable. Indirect quality measures such as the effectiveness of hospital management or the coordination of physician and nurse staff are not directly measurable. However, these indirect measures may affect the aforementioned measurable examples. A tangential hope of this study effort would be that our analysis of the measurable factors leads to a proxy analysis of the non-measurable quality factors. This relies on the assumption that what is measured is a direct or indirect result of these underlying, non-measurable factors.

B. Measuring quality factors

[2] This section includes significant input from Cabe Chadick.

The measurable quality factors as previously discussed are categorical variables, eventually taking on the form of a ranked variable. For example, based on the categorical level of teaching provided in a hospital (none, residency program, medical school affiliation, teaching hospital) an objective 0, 1, 2, or 3 ranking can be given. We can also express a variable that has a defined value (e.g., parking costs in dollar amounts at each hospital) as a ranked variable. Patients rank three hospitals with parking costs of $2, $10, and $30 as the best, middle, and worst facilities (3, 2, and 1, respectively).

Some variables lie between categorical and defined value variables, such as medication adherence rates or mortality rates. These variables can have a defined value for each hospital from 0% to 100%. In practice, we often observe a small number of possible values in hospital reported quality data distributed over a small range. We may rank each hospital for this variable from one to 10 representing non-standard increments (i.e., not 10% increments) in their numerical value. We create these increments based on the distribution of numerical results.

By finding which individual factors in health care correspond best with quality, we hope to be able to quantify a singular hospital-specific factor and ranking. Through systematically prioritizing the factors, top hospitals will not necessarily have the top measure for any of the possible multiple factors. Rather, the best hospital will have the highest combined ranking on the high priority measures.

C. Prioritizing quality factors

The identification of high priority quality determinants is valuable to a number of health care stakeholders, including patients and providers. With such a decentralized health care system in the U.S., consumer decisions drive many of the quality and cost outcomes (Kolstad & Chernew, 2009). Patients may not have the proper clinical background to make health care decisions, and would benefit from effective dissemination of the wide range of available information. Here lies a great need for the prioritization of quality measures. By providing patients a simple subset of ranked measures and hospitals, this format may have the potential to improve decision-making.

Providers need priorities as well, to determine how best to target quality improvement efforts and in measuring progress. Traditionally, providers tend to think in terms of individual process measures rather than view care in terms of the entire process of care or at the population level (Ibrahim, Savitz, Carey, & Wagner, 2001). Finding the most impactful measures can help physicians deliver higher quality care at a population level, rather than just at an individual level.

Actuaries can also benefit immensely from the prioritization process. In this report, we describe the implementation of PRIDIT, a predictive model. We show how to implement PRIDIT "behind the scenes" to rank hospitals. Prioritizing measures of quality facilitates understanding, translating, and communicating the results of the analysis. This can allow actuaries to generate a validated quality score and to communicate to providers, employers, and policymakers the key drivers of those scores. That may make the processes of pay for performance and choosing preferred providers more rigorous, helping health care stakeholders to understand and apply these methodologies.

The PRIDIT method is well suited to accomplishing this prioritization process. PRIDIT can combine many categorical variables, some of which will not be as useful in determining quality (e.g., parking costs, food quality, visiting hours, etc.), some of which are useful proxies (e.g., how often aspirin is administered after a heart attack), and some variables that are outcomes patients really care about (e.g., readmission rate, mortality rate, etc.). PRIDIT works by prioritizing these variables, and then combining them into a single relative measure that correlates with quality. A valid quality score is one that is stable across time and correlated with current or future outcomes measures.

D. Example of the effect of quality measures and their relationships

Quality factors can be important for a number of reasons. They can have a direct, positive impact on patient care. They can be instrumental, as part of a chain leading to the outcomes that patients and providers truly value. They can also be signals of quality with no intrinsic value other than as a guidepost to help patients and providers get to higher quality care. The following is an example of how quality measures relate to one another and how they can be prioritized, ultimately helping us to determine how hospitals compare to one another.

A first pass approach to ranking quality measures

Consider the following three quality factors of a hospital:

- How often the correct body part is operated on
- Infection rates post-op
- Quality of food in terms of how often it is eaten by patients

These three examples illustrate the different types of quality related variables. Operating on the correct body part is an example of a quality measure that has a direct, tangible impact for patients and providers. Both can agree on its importance, and can easily identify adherence to this measure on a case-by-case basis. Infection prevention is a less direct quality measure. Its effects may be observable to the patient, but not directly. Providers will also have some idea about the presence of an infection. Measuring the process of high quality care leading to lower morbidity and mortality shows the value of preventative activities. However, the activities that prevent infection may bring less direct patient satisfaction. Quality of food may be mostly a signal of amenities. While patients may value better tasting food, it will not often be associated directly with higher outcomes. It may be the case that better facilities, or those that command higher rates, provide a better quality "hotel" service to their patients. If such facilities also tend to offer higher quality medical care, then patients and providers could use the quality of the food as a proxy for quality. Without a way to directly experience, measure and observe such quality, signals may be the only way to discern quality. It could also be that, without the information to choose the quality level of their hospital care, patients will make their choices based on the only variables that they know to differ across hospitals, which are amenities and convenience (distance to the hospital). In the absence of evidence or differential payment rates, patients have little cause to make their choice any other way.

It is also important to consider the variance across all of these measures. We presume that the quality of food measure has a tight standard deviation amongst hospitals but the other two, especially the infection rate, have a wide standard deviation. If patients or providers receive feedback about the quality of the food, they can be relatively certain that they will have a similar experience. However, they would need much more data on infection rates and number of safety incidents in order to draw credible conclusions about those aspects of care.

As a result, those studying and utilizing quality information would use this data differently. We would initially give a higher priority to the measure with the widest standard deviation (in this example, infection rates), since those measures would tend to produce variation across hospitals. The hospitals that perform best on that measure are often ranked the best overall. Conversely, we initially would give a lower priority to the measure that has the tightest distribution (here, quality of food). However, those hospitals that did outperform or underperform on tightly distributed measures would be notable as outliers, and we might want to investigate the performance of those specific facilities. This first pass process results in a large number of measures and rankings, without a specific way to combine the evidence collected.

Refining the ranking

Continuing with the example, let us assume that the statistical correlation for the infection rate measure, despite its first-pass high priority, happens to be very low in importance compared to the other measures. Specifically, we observe a measure that has a wide standard deviation as having a small amount of correlation with other measures. In this case, the method downgrades the ranking of that factor, even though it has a wide standard deviation. The theory behind this is that the measure of a hospital's "quality" should be consistent across a broad spectrum of measurable factors. A tangible example is how cheap or expensive hospital parking is. This will vary widely across hospitals (e.g., wide disparity in New York vs. Arkansas hospital parking costs). This may be a factor that has a wide standard deviation and may have cursory "value" to patients, but may not correlate well with other hospital quality factors. In theory, the reverse may also happen, where seemingly unimportant variables can proxy for important, but unobservable, outcomes; all hospital quality measures should be considered when determining overall hospital quality, and this is the primary strength of the PRIDIT method.

E. Introducing the PRIDIT algorithm

The PRIDIT method calculates hospital quality through an aggregation process and is similar to the simple example above. This form of analysis is comprised of two steps. The first step involves calculating a Ridit score for each quality measure.

Motivating Ridit

A Ridit score is a ranking that represents how one hospital ranks relative to the average score for a single quality factor; better hospitals have a better ranking. The score has important mathematical properties that allow us to move from the aforementioned first pass approach to a more refined ranking. Average performance determines Ridit rankings. An example of Ridit ranking would be to give those hospitals with average scores a zero, those with "poor" factor values a negative Ridit score, and those with "good" scores a positive Ridit score. See Table 1 where B and D have average parking costs, C has above average parking cost, and A has below average parking cost. An important requirement of this ranking process is that all Ridit scores must add to zero. This sum-to-zero requirement is important to normalization and principal component analysis, the second step of PRIDIT.

Calculation example

Hospital	Parking cost	Rank	Ridit score
A	\$2	1	0.75
B	\$10	2	0.00
C	\$30	4	−0.75
D	\$10	2	0.00
Average	\$13	2.25	0.00

Table 1: Ridit scores for parking, four hospitals

Hospital	Parking cost	Rank	Ridit score
A	\$2	1	0.80
B	\$10	2	0.00
C	\$30	4	−0.40
D	\$10	2	0.00
E	\$30	4	−0.40
Average	\$16.40	2.60	0.00

Table 2: Ridit scores for parking, five hospitals

Comparing Table 1 with Table 2 shows how the same variable values can generate two different Ridit scores. Hospital A in both tables has the lowest parking cost, which makes it the most preferred by patients on this measure. When the comparator hospitals include three other hospitals (Table 1), with Hospital C as the most expensive at \$30, Hospital A looks best with a positive Ridit score of 0.75. When we add a fourth comparator hospital to the example (Table 2), and Hospitals C and E are the most expensive, Hospital A looks even better, with a positive Ridit score of 0.80. In addition to this, Hospital C has a less negative Ridit score of −0.40, reflecting the fact that it is relatively less of a disliked outlier.

Hospital	Smoking cessation			ACE inhibitor			Proper antibiotic		
	Value	Rank	Ridit score	Value	Rank	Ridit score	Value	Rank	Ridit score
A	90%	2	0.4	99%	2	0.4	100%	1	0.6
B	85%	5	−0.8	92%	4	−0.4	99%	2	−0.2
C	89%	3	−0.2	90%	5	−0.8	98%	3	−0.8
D	100%	1	0.8	100%	1	0.8	100%	1	0.6
E	89%	3	−0.2	93%	3	0.0	99%	2	−0.2
Average	90.60%	2.80	0.00	94.80%	3.00	0.00	99.20%	1.8	0.00
Variance	0.31%	2.20	0.38	0.20%	2.50	0.40	0.01%	0.7	0.36

Table 3: Ridit scores for three quality measures

Hospital	Smoking cessation	ACE inhibitor	Proper antibiotic
A	0.4	0.4	0.6
B	−0.8	−0.4	−0.2
C	−0.2	−0.8	−0.8
D	0.8	0.8	0.6
E	−0.2	0.0	−0.2

Table 4: Final Ridit score matrix

Using a more clinical example, consider Tables 3 and 4 that show three measures reported by Hospitals A through E and their corresponding rankings and Ridit scores. Hospital D gets less credit for 100% adherence to antibiotic guidelines than its 100% adherence to guidelines for smoking cessation counseling and ACE inhibitor usage (see Table 4, row D). The reason for this is that the performance on the proper antibiotic measure is higher on average. Therefore, the comparator group determines that adherence to antibiotic guidelines is less of a proxy for overall quality ranking. This is not a measure of the clinical importance of proper antibiotic use. Rather, the same score on a more difficult to achieve test of quality should count for relatively more in an overall scoring system for hospital quality.

Normalization is the next step, and is a way to rescale the raw Ridit scores. For example, some variables may be binary (yes/no), while others are percentages, and still others are counts. Normalization makes these different types of measures comparable. Normalization is also a way to measure the variation in performance across hospitals. Different measures may have different amount of variation in the data. In this case, the square of the matrix of Ridit scores is the measure of variation. Normalization uses the square root of the sum of square of the overall Ridit score (linear algebra) to rescale the raw Ridit scores. Equation 1 applies normalization to our previous working example:

$$(\text{Ridit matrix})' \cdot (\text{Ridit matrix})$$

$$\begin{pmatrix} 0.4 & -0.8 & -0.2 & 0.8 & -0.2 \\ 0.4 & -0.4 & -0.8 & 0.8 & 0.0 \\ 0.6 & -0.2 & -0.8 & 0.6 & -0.2 \end{pmatrix} \cdot \begin{pmatrix} 0.4 & 0.4 & 0.6 \\ -0.8 & -0.4 & -0.2 \\ -0.2 & -0.8 & -0.8 \\ 0.8 & 0.8 & 0.6 \\ -0.2 & 0.0 & -0.2 \end{pmatrix}$$

Equation 1: Calculation of Ridit score square matrix

	Smoking cessation	ACE inhibitor	Proper antibiotic
Smoking cessation	1.52	1.28	1.08
ACE inhibitor	1.28	1.60	1.44
Proper antibiotic	1.08	1.44	1.44

Table 5: Ridit score square matrix

	Smoking cessation	ACE inhibitor	Proper antibiotic
Square root of diagonal of square matrix	1.23	1.26	1.20

Table 6: Ridit score square root of diagonal of square matrix

Hospital	Smoking cessation	ACE inhibitor	Proper antibiotic
A	0.32	0.32	0.50
B	−0.65	−0.32	−0.17
C	−0.16	−0.63	−0.67
D	0.65	0.63	0.50
E	−0.16	0.00	−0.17

Table 7: Normalized Ridit score matrix

Under normalization, Hospital D gets less credit for 100% adherence to ACE inhibitor use than smoking cessation counseling. The reason is that the ACE inhibitor measure has more dispersion than the smoking cessation measure. Therefore, Hospital D’s outperformance on the ACE inhibitor measure is less of a positive outlier than on the smoking cessation measure. Put another way, 100% performance on the smoking cessation measure is more of a sign of quality because performance is tightly clustered on this measure.

The Principle Component Analysis

The principal component analysis (PCA) compares and combines the normalized Ridit scores to generate a single relative quality measure, which is the final product of our effort. PCA determines the standard deviation of each set of Ridit scores and then gives an initial highest ranking to those factors with the lowest standard deviation of its Ridit scores. PCA then resets this set of factor rankings to give lower rankings to those factors that are not correlated with other factors. Quality measures that have a strong correlation with other measures will keep their higher weightings. PCA generates a final set of factor rankings by iterating this reset process, assessing the variance and covariance of Ridit scores for all factors considered.

PCA yields as many ways of looking at the data as there are variables. This process breaks down the correlations and cross-correlations in the Ridit scores based on components, with the number of components corresponding to the number of variables. Each component explains a portion of the overall variance found in the data. The principle or first component explains the largest degree of variation in the data. The assumption for the case of hospital quality data is that the first component is the component of quality, and that the remaining variation in the data is due to other factors not related to quality. Each component is mathematically unrelated or orthogonal, but the first component being quality is an assumption that we make based on a combination of the data used and our expert judgment.

The first component is the basis for the PRIDIT scores. PRIDIT applies the factor weights to the normalized Ridit scores such that each hospital receives a final weighted average Ridit score, or PRIDIT score. This final ranking is a proxy measure of the underlying latent characteristic of overall hospital quality (the first component). In the example of the three quality measures used, the eigenvalues determined by the PCA process show the relative importance of the first component over the other two (see Table 8).

The meaning of PRIDIT will change based on the type of data used to calculate the scores. The first principle component generated from PCA is the basis of the PRIDIT algorithm, which combines the variation in all variables input into the system. There is no dependent or “left hand side” variable to use to assess the results of PRIDIT. The actual result of PRIDIT is simply the strongest determinant of variation in the data—what we choose to put into the model determines what we get out of it. To use the aforementioned example of amenities (i.e., parking and food), if we simply measured and input amenities, then the PRIDIT score would likely reflect hospital profitability, or the average wealth of patients, or some other factor unrelated to quality. As a result, we review the meaning of individual PRIDIT scores as a proxy for quality.

Component	Eigenvalue
Quality (first component)	2.67
Second	0.29
Third	0.04

Table 8: Eigenvalues for the three components

Measure	PRIDIT weight
Smoking cessation	0.90
ACE inhibitor	0.98
Proper antibiotic	0.95

Table 9: PRIDIT weights for each measure

Given the normalized Ridit matrix, the largest eigenvalue, and the PRIDIT weights, we can calculate the final PRIDIT scores. In this way, we are combining three elements:

1) The normalized Ridit matrix is the performance of each hospital on each quality measure, taking into account relative performance;
2) The PRIDIT weights are the multiplicative terms that represent the variance of each measure and its covariance with all other measures, analyzed via PCA; and
3) The eigenvalue is a scaling factor that puts all scores in the range $(-1,1)$.

Equation 2 shows the final formula for the PRIDIT scores, where the first multiplication is matrix multiplication and the division step involves dividing all values by a scalar.

$$\left(\text{Normalized Ridit matrix}\right)\cdot\left(\text{PRIDIT weights}\right)/\text{ maximum eigenvalue}$$

$$\begin{pmatrix} 0.32 & 0.32 & 0.50 \\ -0.65 & -0.32 & -0.17 \\ -0.16 & -0.63 & -0.67 \\ 0.65 & 0.63 & 0.50 \\ -0.16 & 0.00 & -0.17 \end{pmatrix} \cdot \begin{pmatrix} 0.90 \\ 0.98 \\ 0.95 \end{pmatrix} / 2.67 = \begin{pmatrix} 1.08 \\ -1.06 \\ -1.40 \\ 1.68 \\ -0.31 \end{pmatrix} / 2.67 = \begin{pmatrix} 0.40 \\ -0.40 \\ -0.52 \\ 0.63 \\ -0.12 \end{pmatrix}$$

Equation 2: Calculation of PRIDIT scores

Hospital	PRIDIT score
A	0.40
B	−0.40
C	−0.52
D	0.63
E	−0.12
Average	0.00

Table 10: Final PRIDIT scores

The mathematical properties of the PRIDIT score also guarantee that this is the most efficient use of the available data[3]. Expert judgment determines which variables to include in the analysis, as well as how to code subjective categorical variables. Crucially, the use of the first principal component analysis relies on the assumption that the latent factor that the algorithm identifies is the component representing overall hospital quality. Given that assumption, PRIDIT can utilize a wide array of data and aggregate it into a single score in an efficient manner.

F. Validation of PRIDIT

The novel aspects of the application of PRIDIT in this study are in the aggregation of many different types of quality data, and in the comparison of scores over time. This allows us to validate the method. Aggregation of many types of quality data is now possible because the Hospital Compare

[3] See Brockett, P. L., Derrig, R. A., Golden, L. L., Levine, A. and Alpert, M. (2002), Fraud Classification Using Principal Component Analysis of RIDITs. Journal of Risk and Insurance, 69: 341–371 and Lieberthal, R. D. (2008), Hospital Quality: A *PRIDIT* Approach. Health Services Research, 43: 988–1005 for additional technical details.

database has expanded from its original ten process measures. Additionally, the availability of so much data allows for validation of PRIDIT scores using some data elements against other data elements. Specifically, we can generate a rich set of scores using structural, process, and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) data, and compare the results to mortality outcomes. There is enough data to deal with problems of missing data and to allow us to validate our results.

Comparison of scores over time is possible because Hospital Compare has now been collecting data for long enough that many years of comparable data exist. Thus, scores generated using 2008, 2009, and 2010 data are comparable to those using 2011 data, even though the data collected has changed slightly over time. Thus, PRIDIT allows us to fit the same model to many types of data over time. Comparing scores over time will give us an idea of the stability of the model, while comparing scores to prospective outcomes will demonstrate the value of PRIDIT as a predictive model.

Fitting PRIDIT to the current data set represents a replication of the prior application, as well as new methodologies. In a way, adding additional variables within the measure categories, as well as other measure types, is an exercise in adding additional data to the PRIDIT method, which is not restricted to a certain number of variables. Additionally, the question of variable choice in PRIDIT is not settled. As a result, this study involves adding additional data elements to PRIDIT, and refining the scores. That allows us to move from the mechanical application of PRIDIT to its validation for use in hospital quality improvement.

II. Data

A. Hospital Compare

The main source of data collected for this study comes from the Hospital Compare database, available for download via the Department of Health and Human Services.[4] Hospital Compare focuses on a number of disease states in their reporting; for the purposes of this study, we incorporated measures on heart attack, heart failure, and pneumonia.

---

[4]http://www.hospitalcompare.hhs.gov/

Structural measures indicate the presence of physical equipment or utilities, such as having an emergency department. Hospital Compare also contains data on hospital ownership status, acute care or critical access status, and Veterans Affairs (VA) ownership. Table 16, Table 17, and Table 18 show the counts of different types of hospitals. Additional structural data came from the American Hospital Association as described in section II.B.

Process measures capture the actions performed within the hospital, such as the rate at which the hospital provides aspirin to heart attack patients upon admission. A process measure reflects the actions performed within a system. In the health care system, these measures can represent actions such as smoking cessation counseling and the timing of appropriate antibiotics. Process measures reflect the care a hospital provides to the patient. Hospital Compare collects process measures in the following areas of health care, as shown in Table 19, Table 20, and Table 21:

- AMI: Acute Myocardial Infarction (heart attack)
- HF: Heart Failure
- PN: Pneumonia
- SCIP: Surgical Care Improvement Project

Outcomes measures capture the results of care, such as the mortality rate for a certain disease. Hospital Compare generally collects process measures and HCAHPS data over a 12-month span, so the data we use reflects hospital performance in 2011 and late 2010.The collection period for the mortality and readmission measures is 36 months. Medicare claims and enrollment data are the basis for 30-day risk-adjusted mortality and readmission measures for heart attack, heart failure, and pneumonia, along with the associated patient counts for each measure, as shown in Table 22 and Table 23.

We use patient satisfaction measures obtained through HCAHPS scores in our analysis along with general hospital demographics, structural, process and outcomes measures. HCAHPS is a standardized survey instrument used to measure patients' perspectives in hospital care i.e., patient satisfaction. The measures covered in the survey span a variety of patient experiences. HCAHPS is a form of patient reported measures, in which the source of the measure comes directly from the patient's perspective. Of note: VA hospitals do not report HCAHPS measures to CMS, and thus their satisfaction scores are not included in the analysis.

We also excluded certain HCAHPS measures due to collinearity. These questions required participants to rank their satisfaction with an experience. An example question is, “Did you receive help quickly?” with the choices of Always, Sometimes, or Never. Hospitals report the percentage of patients that chose each selection. The three percentages must add up to 100%. It is thus redundant to include all three HCAHPS measures, since two of the variables are a sufficient statistic for satisfaction on a given measure. We detail collinear HCAHPS in appendix 2 (section IX.A).

In this project, we chose the most negative rating (Sometimes/Never, No and options 0-6 on a scale of 10) as the reference category, and thus excluded it. In interpreting the results, the quality scores are relative to the lowest satisfaction rating. Nineteen variables remained after we stratified the data as shown in Table 24.

B. American Hospital Association

We purchased additional hospital demographic data from the American Hospital Association. This data supplements the data available in the Hospital Compare database. The three types of data we purchased cover teaching status, hospital accreditation, and other hospital characteristics.

Teaching status is important because it has been linked to higher quality in previous studies. However, the relationship is an open question in the literature (Vartak, Ward, & Vaughn, 2008). Teaching status in our study consists of three variables. The first is whether the hospital has a medical school affiliation. The second is whether the hospital has an Accreditation Council for Graduate Medical Education (ACGME) approved Residency program. The final variable is membership in the Council of Teaching Hospitals (COTH), an organization of major teaching hospitals and health systems. The full list of counts of hospital affiliation types is in Table 25.

Accreditation is important because it is frequently an indicator of higher quality. However, the value of accreditation for determining quality is still debated; in at least one study, there was “...considerable variation within accreditation categories in quality of care and mortality among surveyed hospitals...” (Chen, Rathore, Radford, & Krumholz, 2003). Accreditation in our study consists of two variables. These variables represent two hospital accrediting bodies, The Joint Commission and DNV Healthcare. The full list of counts of hospital affiliation types is in Table 26.

Other hospital characteristics include community hospital status, membership in a hospital network, being part of a cluster of hospitals, and number of beds (Table 25). Community hospital status may be important to patients, although the overwhelming majority of hospitals are community hospitals, so this is not likely to be a differentiator of hospitals. Hospital networks and clusters may be important if common action by all hospitals in a network or cluster means that they are of equal quality. Bed size is important because of the positive, demonstrated volume-outcome relationship, which may mean that larger hospitals are better hospitals. Larger hospitals are also more likely to have enough cases to report data to Hospital Compare (see sections II.A and IX.A). Summaries of these elements are in Table 27.

C. Combined data

We combined the Hospital Compare and American Hospital Association data to create a dataset that contained all measures. The advantage of using all available data lies in the ability to judge the hospitals from multiple points of view. Process measures look at whether the hospital is carrying out care in a clinically validated way, while HCAHPS measures look at the patient perspective on care. The outcomes measures, which are risk adjusted, then show the results that the hospital obtains across a number of conditions. There is also a clear monotonic relationship between individual measures and overall quality.

The other major advantage of using all data lies in the crosscutting ability to judge the overall quality of the hospital and the relative usefulness of different types of measures. This result derives from the degree of variability within each measure, and between measures. Presumably, measures of different types have different characteristics. For example, process measures might be inherently more variable than outcomes measures. To take another example, multiple process measures across different conditions might produce meaningful variation in outcomes, while multiple mortality outcomes across different conditions might be similar. Variables that show the greatest weighted values on those measures that are most impactful have the most effect on the relative ranking of hospitals. We also used the full data set to create subsets of the data to allow for analysis of hospitals on specific measure types.

We also created combined data for the years 2010, 2009, and 2008. We used a similar methodology as for the 2011 data, using the Hospital Compare data from each year in combination with the American Hospital Association data from 2011. As a result, we were unable to use the same variables over time, as Hospital Compare adds data elements every year, and drops a small number of variables. We chose 2008 as our cut-off because that was the year Hospital Compare added outcomes data.

III. Analysis

In this section, we describe our analysis of data presented in section II. This analysis generates the results shown in section IV.

A. Fitting current data set

In applying the PRIDIT model to the full data set for 2011, this includes the hospital characteristics, process measures of care, outcome measures of care, and the HCAHPS satisfaction scores. The PRIDIT model generates a single score between −1 and 1 showing hospitals' relative quality. The composite score measures the variance of variables individually and their covariance with other measures as described in section I using Principle Components Analysis. PRIDIT selects the first component that explains the greatest degree of variation observed in the data.

When we apply PRIDIT to the 2011 data, we also generate a score between −1 and 1 showing measures' relative weight in determining PRIDIT score. This score is a relative weight reflecting the statistical importance of each variable, with positive numbers showing a positive association with process and negative numbers showing a negative association with process. Larger numbers in absolute value terms indicate variables of greater importance. For example, if measure A has a weight of 0.8, B has a weight of −0.6, C has a weight of 0.4, and D has a weight of −0.4, then A is relatively more important than B, which is relatively more important than C or D. A is also twice as large as C or D, so it carries twice as much weight—variable weights are linear in scale. Additionally, while C is positively associated with quality, D is negatively associated with quality.

Finally, we analyze the distribution of hospital PRIDIT scores. We calculate the median, mode, and higher moments of PRIDIT scores to show the dispersal of process performance, and create histograms to represent the distribution of hospitals. The mean of PRIDIT is fixed at zero. The median shows whether the fiftieth percentile hospital is relatively high quality (positive) or low quality (negative). The modal range shows where most hospitals are in terms of quality—in Lieberthal, 2008 the median and modal hospitals were slightly below average quality. Since PRIDIT is nonparametric, the standard deviation, skewness, and kurtosis of the distribution of hospitals are freely determined by the data—there could be relatively large dispersal or relatively many hospitals of very high or very low quality.

B. Fitting subsets of the data

We next apply PRIDIT to subsets of the 2011 data. We show the full range of subsets of the data we considered in Table 11. We analyze process measures, outcomes measures, and HCAHPS measures separately, generating PRIDIT scores associated with each measure type. We also analyze the three paired combinations of the variable types: process + outcome, process + satisfaction, and satisfaction + outcomes. All analyses also include hospital characteristic variables. These analyses generate a single score between −1 and 1 showing hospitals' relative performance on the included variable type. We consider this score to be a ranking of the relative success of hospitals when only considering subsets of the full range of measures. Numbers that are more positive indicate higher performance. The composite score measures the variance of the chosen measures and hospital demographic characteristics by themselves and their covariance with other measures within the subset of the data. PRIDIT selects the first component that explains the greatest degree of variation observed in the data set used in each analysis, which is a subset of the overall database.

	Variable types included				
Data set	Structure	Process	Outcomes	Satisfaction	Notes
Process	x	x			
Outcomes	x		x		
Satisfaction	x			x	
Process and outcomes	x	x	x		Results available on request
Process and satisfaction	x	x		x	Results available on request
Outcomes and satisfaction	x		x	x	Results available on request

Table 11: 2011 PRIDIT additional analyses of subsets

When we apply PRIDIT to measure subsets, we generate a score between −1 and 1 showing measures' relative weight in determining PRIDIT score. This score is the relative weight of the statistical importance of the included measures, with positive numbers showing a positive association with process and negative numbers showing a negative association with process. Larger numbers in absolute value terms are also more important variables as with PRIDIT applied to the full data. Finally, we calculate median, mode, and higher moments of PRIDIT scores, and histograms of performance, as with the full data set.

C. Fitting prior sets

We next apply PRIDIT to the 2010, 2009, and 2008 data. These analyses generate a single score between −1 and 1 showing hospitals' relative performance in each year. We consider this score to be a ranking of the relative success of hospitals, with more positive numbers measuring higher performance. We then generate a score between −1 and 1 showing measures' relative weight in determining PRIDIT score in prior years. We consider this score to be a relative weight of the statistical importance of the included measures, with positive numbers showing a positive association with process and negative numbers showing a negative association with process. Larger numbers in absolute value terms are also more important variables as with PRIDIT applied to the full data. Finally, we calculate median, mode, and higher moments of PRIDIT scores, and histograms of performance.

D. Validating PRIDIT

In order to assess this performance across time, we calculate the correlation of 2011 PRIDIT scores with 2010, 2009, and 2008 PRIDIT scores. We also calculate the correlation of variable weights in 2011 with those in 2010, 2009, and 2008. Since the performance of hospitals may fluctuate over time, looking at the PRIDIT results over time will generate insights into the stability of hospital performance. If PRIDIT results show a high degree of correlation across time, or for certain hospitals, then patients, providers, and payers using PRIDIT to make choices for the future can be relatively sure that they are using a stable measure of quality.

The stability of the results is also a way of testing the validity of PRIDIT itself, both for hospital quality and more generally. The quality of hospitals is often stable across time in other studies (Landrigan et al., 2010). PRIDIT should show similar results. Further, if scores are stable across time, then PRIDIT is a predictive model of future hospital quality. Finally, we develop PRIDIT to characterize important underlying latent factors in the data, such as quality. If PRIDIT scores are highly unpredictable across time, then that suggests we are not capturing the same latent factor across time. Inversely, if PRIDIT scores are highly correlated across time, then that suggests that we are capturing the same latent factor, presumably quality, across time. Thus, assessing correlations of scores and weights is a test of the validity of the PRIDIT model.

We also assess the predictive power of the model for outcomes measures using the full data set and subsets of our data. We calculate the correlation between PRIDIT scores from the full data set in 2011 and prior years and the measures of mortality and readmission in 2011. We calculate the correlation between full PRIDIT scores from the years 2008-2011 and the 2011 outcomes data alone. By using the full 2011 PRIDIT scores that include outcomes measures, we create a level of expectation for predicting the scores for past years. We then calculate the correlation between 2011 PRIDIT scores based on subsets of the data, as described in section III.B, with 2011 outcomes. Utilizing data sets for 2011 that only contained a subset of the available measures, as shown in Table 12, we test whether a subset of the hospital quality data we have available is sufficient as a proxy for outcomes. It is also a test of whether improving particular types of quality measures can improve hospital quality. An alternative strategy is to target outcomes measures as a means of quality improvement.

		Variable types included				Outcomes
Data set	Year of data	Structure	Process	Outcomes	Satisfaction	2011 mortality and readmission
All data, 2011	2011	x	x	x	x	x
Process	2011	x	x			x
Satisfaction	2011	x			x	x
Process and Satisfaction	2011	x	x		x	x
All data, 2010	2010	x	x	x	x	x
All data, 2009	2009	x	x	x	x	x
All data, 2008	2008	x	x	x	x	x

Table 12: PRIDIT validation data sets

IV. Results

A. Hospital ranking

Looking at the histogram for overall results in Figure 1 shows that the distribution of hospitals includes a large number of just below average hospitals. Recall that the PRIDIT scores are normed to have a mean of zero, but that the median and other aspects of the distribution are unconstrained. The approximately 800 hospitals with PRIDIT scores in the range $(-0.01,-0.005)$ are not all scored that way due to the same mix of variables. For these hospitals, they report a large number of "N/As" due to low caseloads or not reporting HCAHPS data. They then have demographic variables, tend to report outcomes caseloads, and report a small number of other process, outcomes, and potentially some HCAHPS variables. Hospitals tend to be rated by patients as below average, but not by much, as a result of having lots of missing data, and reporting a small number of outcomes without a strong weight. This is equivalent to the problem reported by Silber et al. (see Silber et al., 2010) with the Medicare Mortality model—when there is a large amount of missing data, hospitals tend to receive similar rankings. PRIDIT distinguishes these hospitals from the average hospital, and ranks them differently, despite the large amount of missing data. This is also why the two-class aspect of PRIDIT is crucial. While a number of hospitals have similar PRIDIT scores in the range $(-0.01,0.005)$, hospitals can still be distinguished as belonging to the high quality or low quality class.

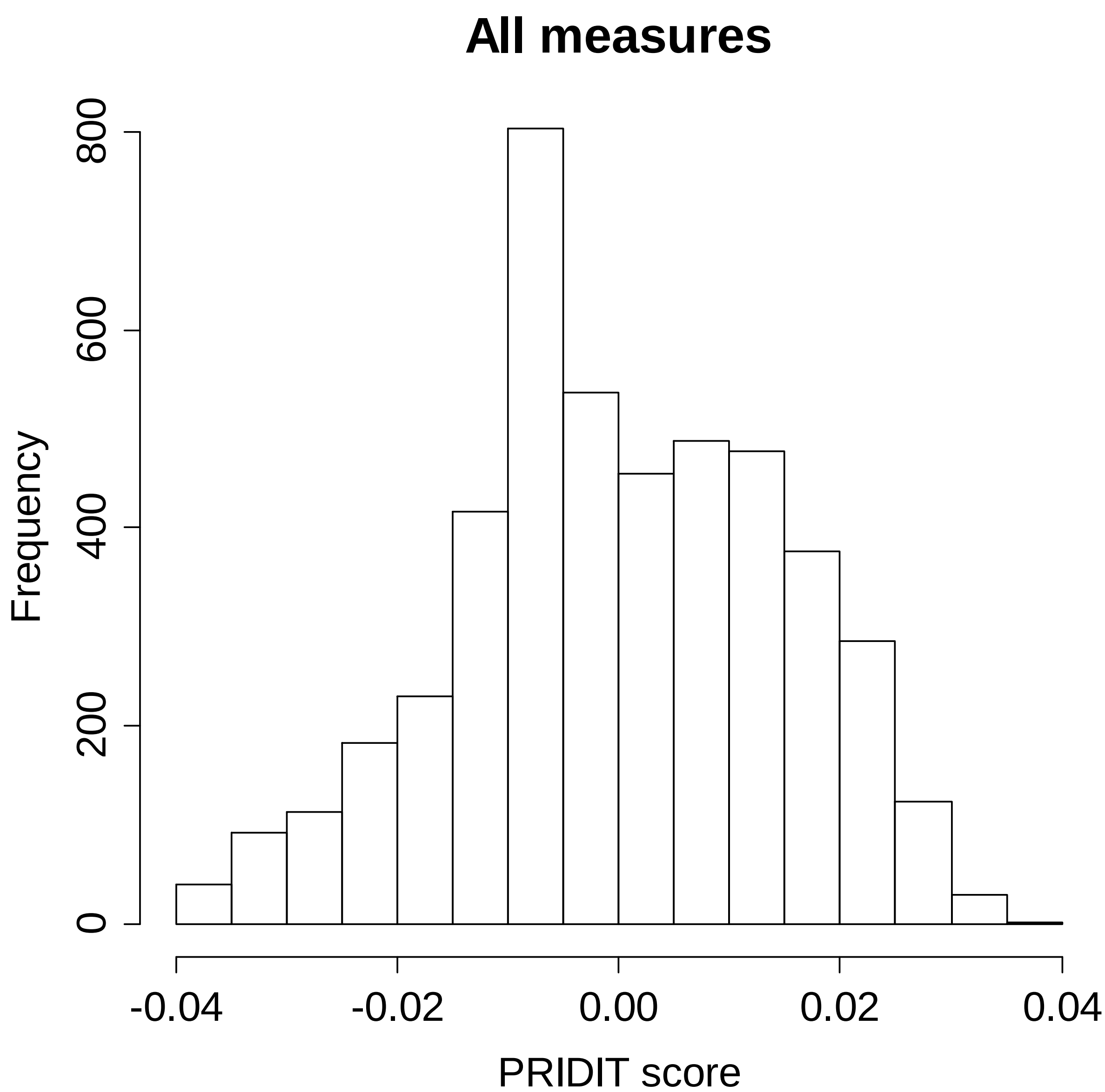


Figure 1: PRIDIT score distribution with all 2011 data

Overall, the dispersal of hospital quality is fairly even, with a slight tendency for hospitals to be worse than average, and a small number of very high and very low quality hospitals. Most hospitals' scores fall into the range $(-0.015, 0.015)$. Most hospitals are of average quality and the median hospital is of just below average quality.

Statistic	Value
Mean	0.00
Median	0.00
Standard deviation	0.01
Skewness	−0.13
Kurtosis	2.54

Table 13: PRIDIT summary statistics

The range of hospital scores $(-0.015, 0.015)$ is also the two standard deviation range of scores. Most hospitals are not that far from the average. However, there are qualitative differences between average, high quality, and low quality hospitals. Smaller, independent, non-teaching hospitals are often on the lower end of the range of scores. These hospitals may not even seek accreditation, although the majority are still accredited. The larger, health system based hospitals with some degree of teaching are on the higher end of the range of scores. They may not have achieved the maximum level on quality indicator—they may have some degree of teaching or some level of integration with a hospital system—but they have enough of these characteristics, and other process measures, outcomes, and consumer satisfaction scores, to rank as higher quality hospitals in PRIDIT.

The higher moments of skewness and kurtosis show additional attributes of the distribution of scores. The skewness does reflect the fact that hospital quality distribution skews towards the negative, as there are more below average than above average hospitals. However, the distribution of hospitals is fairly even—the minimum and maximum scores are similar values with opposite signs. The kurtosis also demonstrates the number of extreme valued PRIDIT scores—a Normal distribution has a kurtosis of 3, so this distribution has thinner sized tails i.e., less extreme valued scores than the Normal distribution.

A notable aspect of the histogram is the possible bimodal distribution in the middle of the PRIDIT distribution. In the middle of the graph, the bars go down and then up again. This may indicate that the higher and lower quality hospitals are of two different types. It may also mean that we should split the sample, with better and worse hospitals analyzed separately. For example, in analyzing the height of a population, it would be important to split the sample between men and women. Otherwise, the main factor driving the distribution of heights would be gender, and that could come out in the graph as a bimodal distribution of height in the population (Iversen & Gergen, 1997; Schilling, Watkins, & Watkins, 2002). If the multifactorial nature of quality were different for different types of general, acute care hospitals that were as distinct as men and women, then the correct procedure would be to split the sample and analyze each type of hospital separately. While that analysis is beyond the scope of this report, it is relevant to note that, for the hospitals in general, their scores are continuous. In other words, the number of hospitals in each bar in the histogram goes up and then down. This indicates that hospitals can be grouped together and scored against one another. In other words, we do not need separate methods to score academic medical centers, rural community hospitals, VA hospitals, and so on. That is for the positive since the more comparator hospitals, the better the data is, and the more potential there is for lessons learned to apply to hospital quality improvement.

We characterize the top 100 hospitals in Table 14 and Figure 2 highlights that the distribution of high quality hospitals is not spread evenly across the country. The top hospitals are concentrated in the New York—Washington DC corridor, and the upper Midwest, with a smaller number in Florida and California. There could be a number of explanations for this finding. One is the preponderance of academic medical centers in the northeast. However, this explanation cannot be the entirety of the story. The reason is that, looking at the value of variable weightings shown in Table 28, the residency, medical school, and teaching hospital membership variables are not in the top 20 of variable rankings. Thus, the observed results should reflect a larger number of factors. The fact that the hospitals are geographically clustered likely reflects the fact that the PRIDIT system is uncovering a common latent characteristic in the data. The more data that is used in the analysis, the better we are able to sift out the top performing hospitals; the preponderance of data is also necessary as the concept of quality itself is multifactorial and needs to be measured through a number of variables.

	Population		Top 100 hospitals		
Region	Count (million)	Percent	Count	Percent	Density (per million)
Northeast	55	18%	50	50%	0.91
Midwest	67	22%	23	23%	0.34
South	115	37%	16	16%	0.14
West	72	23%	11	11%	0.15
United States	309		100		0.32

Table 14: Top 100 hospitals by region

Figure 2: Location of top 100 hospitals

B. Variable importance

The full variable ranking is in Table 28; we highlight the top ten measures in terms of the absolute value of their weights. One significant finding is that all of the top ten measures of quality are consumer satisfaction scores from HCAHPS measures. Also of note is that the highest consumer satisfaction scores are negatively associated with quality. Mid-level scores are positively associated with quality, while the lowest category is the reference category that we excluded from the analysis. In many cases, the score for the middle category is of similar magnitude, but in the opposite direction, similar to the top measure. The implications are twofold. First, the best hospitals are not the ones that are the quietest or that have the most responsive clinicians. Busy hospitals tend to have better performance, which is consistent with the volume-outcome relationship (Luft, Hunt, & Maerki, 1987). These hospitals score highly on process measures of care and outcomes, especially volume, but only in the middle in terms of consumer satisfaction. PRIDIT interprets that pattern of correlation by relating high quality to the highest scores on process measures and outcomes measures, and mid-level scores on satisfaction. That is one explanation for the pattern of variable weights generated by PRIDIT. Secondly, the PRIDIT method is reducing the value of overall consumer satisfaction by weighting these two measures in the opposite direction. PRIDIT observes the high degree of variation in top achievement in satisfaction and mid-level achievement in satisfaction, and thus ascribes to each a strong, opposite signed, variable weight. Taken together, these variable weights largely cancel out. Thus, the contribution of individual weighting of satisfaction variables to PRIDIT scores is less than the individual variable ranks imply.

The weighting also shows that while mortality rates are negatively associated with quality, readmission rates are positively associated with quality. The weights on both types of measures are relatively small when compared to total patient counts, i.e., the number of patients with a particular condition (AMI, heart failure, or pneumonia) eligible for measurement on a particular outcome (mortality or readmission). This may occur because mortality and readmissions are risk adjusted and these measures do not vary much. Additionally, readmissions are not necessarily a signal of poor quality, which may argue against programs to penalize hospitals for high readmission rates (Axon & Williams, 2011). Thus, when assessing hospital quality, risk adjusted outcomes are less informative than volumes. The characteristics variable “number of beds” shows the same important relationship, where larger hospitals have higher quality in addition to the outcome count variables.

When interpreting results, it is important to note the difference between statistical and clinical significance. In our case, the process measures of care drop out in terms of importance; none of them appears in the top 20 measures based on performance. Although the PRIDIT ranking process may not rank these measures as high when calculating overall quality, these process measures (specifically measures that are derived from evidence-based clinical care guidelines) are crucial to the delivery of care and are no less important clinically than other measures that are ranked higher. The ranking of these measures is probably due to their small amount of variation and reflects the low marginal value of these variables. Once we add outcomes, satisfaction, and characteristics into the PRIDIT model, the process measures may add little additional information. That might be evidence for putting fewer resources into collecting them, or even not using such variables for pay for performance programs.

C. Validation of PRIDIT

The value of PRIDIT scores using the full data is highly correlated across time. The correlation coefficient for the 2010 and 2011 PRIDIT scores is 0.93. This finding is consistent with the hospital quality literature, where the performance of hospitals across time is both highly stable and difficult to impact with quality improvement or pay for performance programs (Landrigan et al., 2010). This is a positive result for the use of PRIDIT, since the scores in a given year should be highly predictive of future performance in general. A single year’s PRIDIT score can be used to contract over multiple years. One aspect of the data that biases the correlation upwards is the fact that the correlation is only available for those hospitals that reported data in both years. Since the number not available for 2010 is small (98), this should not be a major driver of the results.

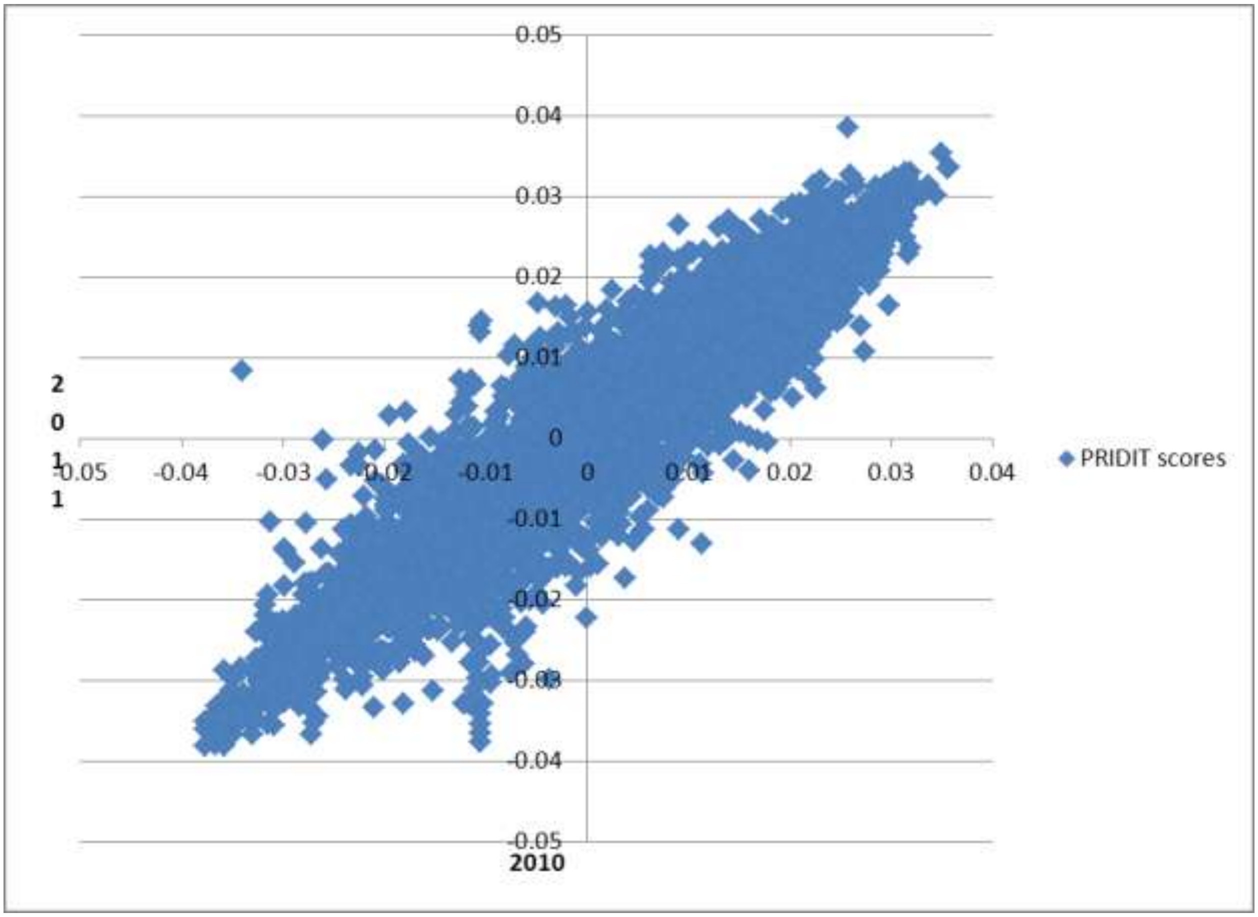
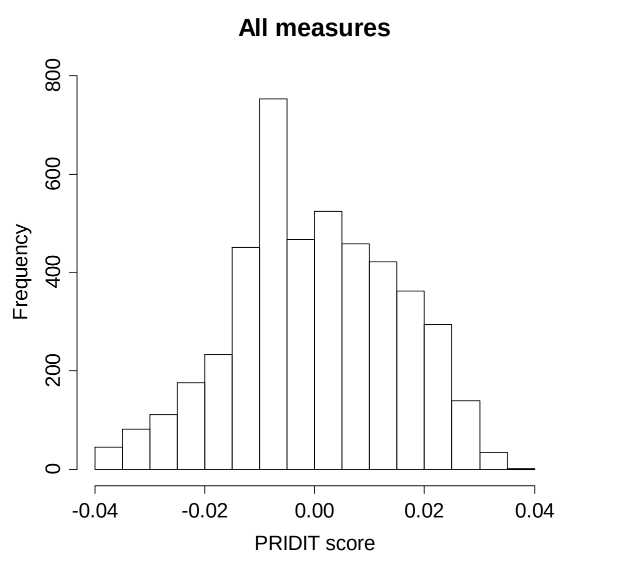


Figure 3: Correlation of PRIDIT scores across time

In our comparison of hospital variation in 2010 and 2011, there was also a high degree of consistency. There are a large number of slightly below average hospitals in both years. There are also a small number of extremely high and extremely low quality hospitals. There is also a bimodal distribution in both years. However, in 2011, the large mass of hospitals is worse i.e., farther from average than in 2010. Thus, the lower quality hospitals are easier to distinguish from the average in 2011 than in 2010 (see Figure 4 and Figure 5).



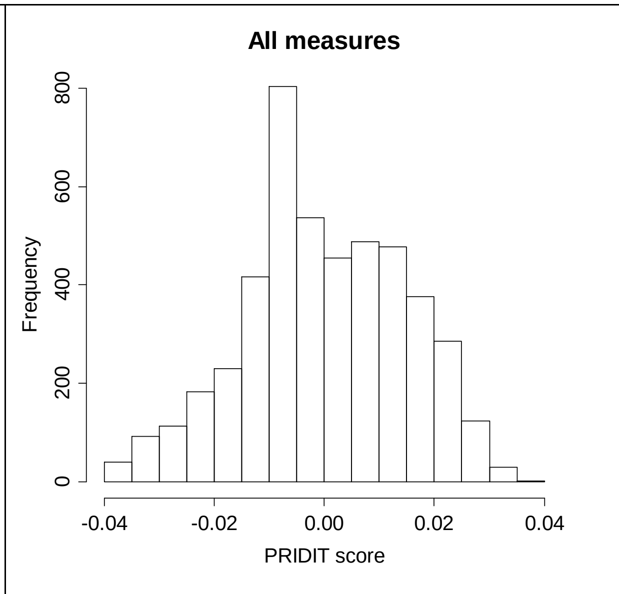


Figure 4: 2010 PRIDIT scores with all data	**Figure 5: 2011 PRIDIT scores with all data**

Another outcome of interest is the correlation of PRIDIT scores and outcomes measures. We measured this variation two ways. First, we looked at the correlation between scores derived from structural, process, and satisfaction measures with outcomes of mortality and readmission. We also looked at the correlation of 2010 PRIDIT scores with 2011 outcomes, to determine the predictive value of PRIDIT.

The correlation between the full PRIDIT results including all outcomes for 2011 and the outcomes of mortality is consistent with prior work in the area of hospital quality. The correlations of scores and deaths due to heart attack, heart failure, and pneumonia are −0.17, −0.19, and −0.09, respectively. In other words, higher PRIDIT scores are correlated with lower risk adjusted mortality rates. The degree of the correlation is relatively low, which is consistent with evidence that hospital mortality may be difficult to explain or to change (Ryan, Nallamothu, & Dimick, 2012). The correlations of scores and readmissions due to heart attack, heart failure, and pneumonia are positive: 0.12, 0.12, and 0.19. This result is consistent with the relatively small weighting given to these measures when they are included in the PRIDIT score, and the fact that mortality is negatively weighted for PRIDIT while readmissions is positively weighted for PRIDIT.

The use of only process measures with structural data shows a consistent but less highly correlated view of outcomes. The correlations of structural and process scores with heart attack, heart failure, and pneumonia death are all −0.09. The correlations of structural and process scores with heart attack, heart failure, and pneumonia readmissions are essentially zero, −0.01, −0.02, 0.02, respectively. Thus, hospitals that have strong process measures can expect to have lower mortality. In fact, for pneumonia, adding mortality rates to PRIDIT does not improve the correlation, showing that we could judge hospitals on process measures alone. For readmissions, it is a different story—process measures seem to have no bearing on risk adjusted readmissions rates.

The use of only HCAHPS measures along with structural data shows a very different view of outcomes. The correlations of structural and HCAHPS scores with heart attack, heart failure, and pneumonia death are 0.08, 0.15, and 0.04, respectively. Higher satisfaction is associated with higher mortality rates. The correlations of structural and HCAHPS scores with heart attack, heart failure, and pneumonia readmissions are −0.12, −0.13, and −0.16, respectively. Higher satisfaction is strongly correlated with a lower likelihood of risk-adjusted readmission. Such hospitals also have much lower volumes, with correlations with the patient count measures in the range $(-0.34, -0.30)$. High satisfaction hospitals have lower volumes, lower readmissions, and worse mortality. The latter relationship is consistent with the volume-outcome principle. The results are broadly similar when we added process measures of care, showing that satisfaction variables dominate process measures of care variables in the PRIDIT results.

	HA Death Rate	HA Death N	HA Read Rate	HA Read N	HF Death Rate	HF Death N	HF Read Rate	HF Read N	PN Death Rate	PN Death N	PN Read Rate	PN Read N
2010	−0.19	0.53	0.12	0.47	−0.20	0.58	0.10	0.57	−0.11	0.57	0.17	0.56
2009	−0.18	0.53	0.11	0.47	−0.20	0.58	0.08	0.57	−0.11	0.56	0.15	0.55
2008	−0.09	0.30	0.11	0.26	−0.14	0.34	0.11	0.34	−0.05	0.30	0.13	0.30

Table 15: Prior PRIDIT scores and 2011 outcomes

Finally, we look at the strong correlation between 2010 full PRIDIT scores and 2011 outcomes. The correlations of structural and process scores and heart attack, heart failure, and pneumonia death are −0.19, −0.20, and −0.11, respectively, as shown in Table 15. 2010 PRIDIT scores are more correlated with 2011 outcomes than the 2011 scores. This may be due to the aforementioned selection effect, with those that reported in 2010 and 2011 more likely to have consistent outcomes. The correlations with heart attack, heart failure, and pneumonia readmissions are 0.12, 0.10, and 0.17, respectively. These are similar, but smaller than, the correlation between full 2011 PRIDIT scores and 2011 readmission rates. We found a similar degree of correlation between 2009 PRIDIT scores and 2011 outcomes, and smaller correlations between 2008 PRIDIT scores and 2011 outcomes. These results all suggest that PRIDIT is a valid predictive model for hospital outcomes. It takes three years of data for the predictive power of PRIDIT to decline significantly.

As demonstrated in Figure 6, the variable weights are highly correlated over time. In fact, the 2010 and 2011 weights are perfectly correlated (correlation coefficient > 0.99). This is another demonstration of the stability of PRIDIT for hospitals over time, this time in the weights. The first component explaining the variances and covariances of variables are highly stable over time. The stability exists despite the fact that there are two variables in the 2011 data that are not in the 2010 data (ER and SCIP_INF_9).

The figure also shows the types of variable weightings for determining quality. There are many variables clustered around zero. PRIDIT positively weights many variables for quality. PRIDIT weights fewer variables negatively for quality, such as mortality rates and highest levels of satisfaction. The point is that we can choose a small subset of the variables in order to measure quality efficiently, because we can remove those variables that have weights close to zero. Additionally, the full set of variables in one year, or certain variable subsets we choose, will be valid in the next year.

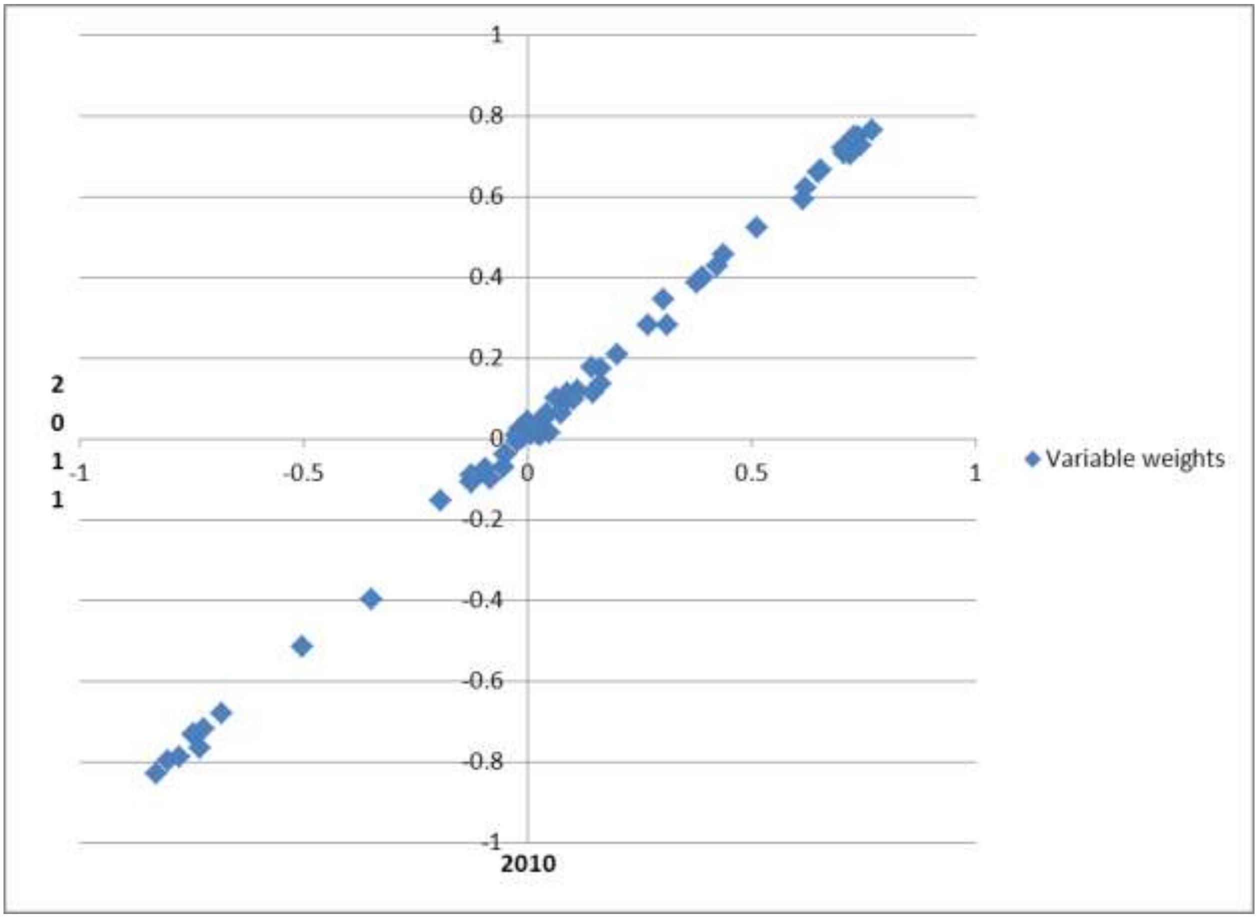


Figure 6: 2010-2011 variable weight comparison

V. Conclusions

A. Combining different types of data

Combining different types of data both shows great promise for hospital quality improvement and shows how informative certain measures are. Hospital characteristics and structural measures are an easily observable way to look for signals of quality. Process measures of care are useful indicators that are often collected during the course of care. Outcomes measures cover what patients care about and wish to avoid—mortality and readmission. Consumer satisfaction measures stated patient experience in a hospital.

Combining these types of data also shows that they should not be viewed in isolation. While hospital characteristics are important for measuring quality, they are difficult or impossible for a hospital to change. High readmissions rates may not be a sign of poor quality, even though patients do not like them and Medicare and private payers are focusing on them as a source of cost reduction. Patients do not seem to see their experience as multidimensional, but rather answer all questions based on a single internal measure of how their care went. As a result, consumer satisfaction scores tend to reflect a limited perspective. The use of objective data is critical to provide a holistic view of hospital quality.

B. PRIDIT combined

The meaning of PRIDIT when using all the data is to give an overall picture of hospital quality. This overall picture will include all of the elements that we input into it—structural, process, outcome, and satisfaction measures. How well hospitals score on the variables they report will determine the quality of the hospital, especially those variables with the strongest weight. Variables that are individually important, good indicators of performance within a measure type, and good indicators of performance across measure types will tend to get the highest weights. Similarly, structural characteristics, such as not-for-profit status, affect multiple types of performance. Thus, the PRIDIT score will give as broad a view of hospital performance as is possible.

The other use for overall PRIDIT scores is to judge the distribution of hospitals. The advantage of looking at the distribution is that it is less sensitive to the scores or patient counts of individual hospitals. The distribution can show how many or few truly great hospitals there are, as well as outliers that researchers identify and study for further lessons learned. It can show whether the distribution is continuous or whether there are certain specific quality “types,” such as high, average, and low quality, that are distinct from each other. It can also show over time whether the distribution of hospitals is changing and how.

The combination of data through PRIDIT shows the possibility for prioritizing quality measures, both at a single point in time and across time. At a single point in time, many of the measures we used had little or no effect on quality scores. Hospitals should focus on collecting the measures that will be most useful. These will also tend to be the most useful measures over time. While there may be scope to replace measures as they become less useful, the process of continually adding more measures to Hospital Compare may not be improving quality. Adding a select number of hospital characteristics to Hospital Compare may have a positive impact on quality.

A possible disadvantage of the mixing of data types is that there may be much more variation in one type of data. Then, based on the PRIDIT methodology, that type of data would tend to “dominate” the PRIDIT results. A good example of that is in the case of outcomes data, where the counts dominate the mortality and readmissions rates due to a much higher degree of variation. The difficulty of combining so many types makes this an empirical question—does one form of data dominate the scoring? Otherwise, there is no natural way to set one data as more important than another is. Rather, the degree of variation exhibited by different types of data is an important finding of this study, showing the most useful proxies for hospital quality out of a large set of possibilities.

C. Actuarial practice

The overall quality of hospitals is highly variable across the U.S. Hospital quality varies between acute care and critical access hospitals, by geography, teaching status, and other observable hospital characteristics. Certain patterns of behavior characterize hospital quality. The best hospitals stand out based on high volume, outperformance on certain measures of care, and moderate but not outstanding patient satisfaction. From this point of view, patients truly may not know what type of setting has the highest quality of care, or they may be rationally trading off quality in order to be more satisfied. One of the main roles for actuaries, who may have specialized local knowledge of the markets they analyze, could be to design products that include quality of care. After all, health care is local (Wennberg & Gittelsohn, 1973).

Actuarial implementation of PRIDIT could allow actuaries to differentiate quality of hospitals within a market or between markets for hospital care. It allows for benchmarking of hospitals, locally and nationally, on quality scores. Insurance companies that now have pay for performance programs that collect similar data from hospitals could pay for improvements in PRIDIT scores rather than individual measures, or use PRIDIT scores to pay for outperforming hospitals only. Hospitals might also benefit from comparing their PRIDIT scores based on data from each insurance company, and government payers, in order to target quality improvement efforts.

One application for quality data may be in pricing. This question arose because of the research project, and we consider it an open question. Market dynamics, rather than hospital quality or outcomes, may determine provider reimbursement. It could be the case that higher quality care is costlier. In that case, individuals living in an area with better hospitals would receive better care and thus be costlier to insure. The opposite could also be true—it has been posited that higher quality care is less costly (Solberg, Lyles, Shore, Lemke, & Weiner, 2002). We plan to explore this relationship in future work.

Another insurance application is preferred provider networks. This should be a direct consequence of having different quality levels in a community large enough to support multiple providers. An insurer could select a level of quality that it wishes to associate with a particular health insurance plan. Then, it could selectively contract with the providers at that level. In fact, such an insurer could even offer plans with different quality levels, much as they now offer plans with different deductibles or different designs. PRIDIT gives an objective way to set up, and adjust, such levels over time.

D. Improving hospital quality

Our results show that there is a lot of variation in care in regions with many providers. In these regions, improving hospital quality could involve emulating the highest quality hospital, or even individuals or insurers shifting their business such that lower quality providers consolidate or close. There is also room for additional travel to receive high quality care, especially in the Northeast. The map in Figure 2 shows the advantages this area represents. Patients may not see traveling 50 miles as a choice, but it may lead to higher quality care.

In low-density areas, improving the single hospital available is the way to ensure quality. Some individuals are fortunate enough to live near an “island of quality.” Others live near one of the lower performing hospitals that report little or no quality data, and thus are particularly difficult to identify as low quality without the use of PRIDIT. For such areas, building a new facility is likely not an option. Rather, the focus must be on improving the hospital by focusing on the most impactful measures as identified by hospital quality research. In the extreme, individuals or insurers shifting their business such that the lower quality hospital closes may be an option, as it would result in higher volumes. The PRIDIT method shows several ways to improve quality, and thus provides options for quality improvement efforts.

E. Future work

Future applications of PRIDIT include utilizing the methodology presented in this report to improve other areas of health care quality. The ideal would be to use the same variables and same methodology in order to assess quality in the outpatient, pharmacy, and home care setting. A comparison of which factors and drivers of quality are common across health care settings would be an ideal outcome of future work. In reality, it is difficult to judge different health care settings using the same set of measures, and so different sets of measures have been developed for different types of care. PRIDIT could be applied in the same way in different settings, with potential to use the same outcome measures. To the extent that there are common variables or measures in different settings, such as patient satisfaction, PRIDIT could illustrate the relative importance of the same variables or variable domains in different health care settings. As a result, PRIDIT has the potential to show whether the same types of providers should be measured and rewarded differently in different settings.

A second future application of PRIDIT is fraud detection in health care. The use of PRIDIT for the detection of fraud in other insurance lines such as automobile insurance has been studied. The use of PRIDIT for health care fraud research has been limited or non-existent, as far as the published literature suggests. The potential for applying PRIDIT for health care fraud comes from two sources. One is the basic framework of PRIDIT, which is ideally suited for fraud. The second is the fact that quality varies so widely across providers, and standard of care is much more disputed in health care than in, say, automobile insurance, where specific factory guidelines exist as to vehicle performance and maintenance. Using PRIDIT to merge variables that are potentially indicative of fraud with variables such as quality, outcomes, and satisfaction could generate important new findings in this area. The fact that PRIDIT is highly valid in the study of hospital quality should make it ideally suited for fraud detection, too. In fraud, we might expect to observe less stability in results across years as the perpetrators of fraud learn, and adapt to, fraud detection scores. Whether this is the case is an empirical question that should be tested using PRIDIT as an ideal methodology.

VI. Bibliography

Ai, J., Brockett, P. L., Golden, L. L., & Guillén, M. (2012). A robust unsupervised method for fraud rate estimation. *Journal of Risk and Insurance*. doi: 10.1111/j.1539-6975.2012.01467.x.

Axon, R. N., & Williams, M. V. (2011). Hospital readmission as an accountability measure. *Journal of the American Medical Association, 305*(5), 504.

Bhattacharyya, G. K., & Johnson, R. A. (1977). *Statistical concepts and methods*. New York: Wiley.

Brockett, P. L., Derrig, R. A., Golden, L. L., Levine, A., & Alpert, M. (2002). Fraud classification using principal component analysis of RIDITs. *Journal of Risk and Insurance, 69*(3), 341-371. doi: 10.1111/1539-6975.00027

Brook, R. H., McGlynn, E. A., & Shekelle, P. G. (2000). Defining and measuring quality of care: A perspective from US researchers. *International Journal for Quality in Health Care, 12*(4), 281-295.

Centers for Medicare & Medicaid Services. (2008). *Approval of deeming authority of Det Norske Veritas Healthcare, Inc. for hospitals.* Retrieved September 6, 2012, from http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/SurveyCertificationGenInfo/downloads/SCLetter09-02.pdf

Chen, J., Rathore, S. S., Radford, M. J., & Krumholz, H. M. (2003). JCAHO accreditation and quality of care for acute myocardial infarction. *Health Affairs (Project Hope), 22*(2), 243-254.

Chen, T. T., Lai, M. S., Lin, I. C., & Chung, K. P. (2012). Exploring and comparing the characteristics of nonlatent and latent composite scores implications for pay-for-performance incentive design. *Medical Decision Making, 32*(1), 132-144.

Geweke, J., Gowrisankaran, G., & Town, R. J. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica, 71*(4), 1215-1238.

Ibrahim, M. A., Savitz, L. A., Carey, T. S., & Wagner, E. H. (2001). Population-based health principles in medical and public health practice. *Journal of Public Health Management and Practice, 7*(3), 75-81.

Iversen, G. R., & Gergen, M. M. (1997). *Statistics: The conceptual approach*. New York: Springer.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (3rd ed.). Pacific Grove: Duxbury Press.

Kolstad, J. T., & Chernew, M. E. (2009). Quality and consumer decision making in the market for health insurance and health care services. *Medical Care Research and Review, 66*(1 Suppl), 28S-52S. doi: 10.1177/1077558708325887

Landrigan, C. P., Parry, G. J., Bones, C. B., Hackbarth, A. D., Goldmann, D. A., & Sharek, P. J. (2010). Temporal trends in rates of patient harm resulting from medical care. *New England Journal of Medicine, 363*(22), 2124-2134.

Lieberthal, R. D. (2008). Hospital quality: A PRIDIT approach. *Health Services Research, 43*(3), 988-1005. doi: 10.1111/j.1475-6773.2007.00821.x

Lovaglio, P. G. (2012). Benchmarking strategies for measuring the quality of healthcare: Problems and prospects. *The Scientific World Journal, 2012*(606154), 1-13. doi: 10.1100/2012/606154

Luft, H. S., Hunt, S. S., & Maerki, S. C. (1987). The volume-outcome relationship: Practice-makes-perfect or selective-referral patterns? *Health Services Research, 22*(2), 157-182.

Ryan, A. M., Nallamothu, B. K., & Dimick, J. B. (2012). Medicare's public reporting initiative on hospital quality had modest or no impact on mortality from three key conditions. *Health Affairs (Project Hope), 31*(3), 585-592. doi: 10.1377/hlthaff.2011.0719

Schilling, M. F., Watkins, A. E., & Watkins, W. (2002). Is human height bimodal? *The American Statistician, 56*(3), 223-229.

Shahian, D. M., Wolf, R. E., Iezzoni, L. I., Kirle, L., & Normand, S. L. (2010). Variability in the measurement of hospital-wide mortality rates. *New England Journal of Medicine, 363*(26), 2530-2539. doi: 10.1056/NEJMsa1006396

Silber, J. H., Rosenbaum, P. R., Brachet, T. J., Ross, R. N., Bressler, L. J., Even-Shoshan, O., Volpp, K. G. (2010). The Hospital Compare mortality model and the volume-outcome relationship. *Health Services Research, 45*(5 Pt 1), 1148-1167. doi: 10.1111/j.1475-6773.2010.01130.x; 10.1111/j.1475-6773.2010.01130.x

Solberg, L. I., Lyles, C. A., Shore, A. D., Lemke, K. W., & Weiner, J. P. (2002). Is quality free? The relationship between cost and quality across 18 provider groups. *American Journal of Managed Care, 8*(5), 413-422.

The Leapfrog Group. (2010). *What's new in 2010: The Leapfrog hospital survey.* Retrieved September 4, 2012, from http://www.leapfroggroup.org/media/file/2010_Leapfrog_Hospital_Survey_Overview_TownHallCalls.ppt

U.S. Department of Health and Human Services. *Medicare Hospital Compare information for professionals: Technical appendix.* Retrieved July 12, 2012, from http://www.hospitalcompare.hhs.gov/staticpages/for-professionals/poc/technical-appendix.aspx

Vartak, S., Ward, M. M., & Vaughn, T. E. (2008). Do postoperative complications vary by hospital teaching status? *Medical Care, 46*(1), 25.

Wennberg, J., & Gittelsohn, A. (1973). Small area variations in health care delivery. *Science, 182*(4117), 1102-1108.

Werner, R. M., & Bradlow, E. T. (2006). Relationship between Medicare's Hospital Compare performance measures and mortality rates. *Journal of the American Medical Association, 296*(22), 2694-2702. doi: 296/22/2694 [pii]; 10.1001/jama.296.22.2694 [doi]

VII. Glossary

Collinearity—A measure of variable similarity. A situation in fitting a model to multiple variables when "...one of the predictors (variables) is an *exact* linear combination of the others." Near collinearity occurs when the correlation "...of one predictor with the remaining predictors is nearly 1."(Kleinbaum, Kupper, Muller, & Nizam, 1998, pp. 240-241)

Dispersion—A measure of variability or spread in a set of data. Variability is measured as "...deviations from the mean", which shows how far data points are from the mean, or center, of the data (Bhattacharyya & Johnson, 1977, p. 32).

Eigenvalues—A measure of the importance of the components of a matrix. The variance of the components generated by a principal components analysis (Kleinbaum et al., 1998, p. 243).

Hospital accrediting body—An organization that assesses the overall quality of a hospital. Accrediting organizations generally gain formal recognition "...as a national accreditation program for hospitals seeking to participate in the Medicare program." Currently, the Joint Commission, the American Osteopathic Association, and DNV Healthcare are the three organizations with such recognition (Centers for Medicare & Medicaid Services, 2008).

Kurtosis—A measure of dispersion. It indicates "...the heaviness of the tails relative to the middle of the distribution." (Bhattacharyya & Johnson, 1977, p. 219)

Matrix—A "...rectangular array of numbers." A matrix has rows and columns (Bhattacharyya & Johnson, 1977, p. 732). One example would be the Final Ridit score matrix found in Table 4 of the report, which has five rows and three columns.

Skewness—A measure of dispersion. It "...indicates the degree of asymmetry of a distribution." (Bhattacharyya & Johnson, 1977, p. 219)

VIII. Appendix 1—theoretical model

A. Stating the problem

Let us suppose that there is an unobserved, latent measure of relative quality Q. This measure is ordinal and a scalar, in that for any two hospitals i and j, $Q_i>Q_j$ is equivalent to the statement that hospital i is of higher quality than hospital j. If Q is real valued, then it is possible to rescale Q onto the interval $(-1,1)$ without loss of information.

Let us also suppose that there is a vector of quality measures **q** with n elements. Each member of **q**, q_1, q_2,...,q_n, is a scalar proxy for quality. That is, the correlation of the kth measure of quality with overall quality, corr(q_k,Q) is in the range $(-1,1)$. This measure is ordinal and monotonic, in that for any two hospitals i and j, $q_{k,i}>q_{k,j}$ implies that, all else equal, $Q_i>Q_j$. However, we do not observe Q_i and Q_j. Further, there is no observable proxy $Q^{\dagger}$ for Q. Thus, we wish to find some way to create a proxy $Q^{*}$for Q using the observable vector of quality measures **q**.

B. Application of PRIDIT to solve the quality problem

Given this setup, Brockett et al., 2002, show that there is a method that produces Q_j^{*}, a single number that represents the latent variable Q_i. This measure is the most efficient way to combine the many scalar proxies for quality q, and produces a number scaled to the range $(-1,1)$. The closer a number is to −1, the worse the quality is, and the closer it is to 1, the better the quality is. The average hospital is normed to 0, so that negative scored hospitals are worse than average, and positive scored hospitals are better than average. The scale is also multiplicative, so a hospital with a score of 0.50 is twice as good, in terms of quality, as that with a score of 0.25. On an absolute value basis, the scale is also multiplicative with negative values. A positive hospital score indicates that a hospital is in the “high quality hospital” class, while a negative score indicates that the hospital is in the “low quality hospital” class. Additionally, a hospital score of 0.50 is twice as strong as a score of −0.25, although one indicates high quality class membership and the other indicates low quality class membership. The method produces this score by generating weights **w**, i.e., w_1, w_2,...,w_n for each measure q_1, q_2,...,q_n. These weights are also scaled to the range $(-1,1)$ and multiplicative. While this description of the PRIDIT method applies to hospitals, it applies equally to other applications, such as fraud, that have been described in other contexts (Ai, Brockett, Golden, & Guillen, 2012).

The application of PRIDIT to the problem of hospital quality detection has been previously described (Chen, Lai, Lin, & Chung, 2012; Lieberthal, 2008). These prior analyses have focused on the use of process measures alone to measure quality. We expand on the application to quality, with several additional features.

First, this analysis uses multiple types of measures. The theory does not restrict the use of PRIDIT to a single type of quality data. However, the assumption is that the first principal component represented quality. In mixing different types of data, we explicitly state this assumption.

The second difference is the use of outcomes as a validation measure. One problem with unsupervised learning methods is that there is no standard ($Q^{\dagger}$) to validate quality. We show how to use outcomes to validate the scores obtained from other measures. We also show how to use outcomes as part of the set of variables we analyze with PRIDIT.

Finally, this analysis uses PRIDIT scores over time to capture the stability of quality measures. Hospital scores should be consistent over time as long as hospital performance does not change radically. Variable scores should also be consistent over time. This is because we are aggregating many proxies to measure a single latent measure, and that measure belongs to the entire institution. The analysis will show the exact degree of consistency over time.

C. Theoretical motivation for measure inclusion

Process measures are a validated way to measure the quality of how hospitals perform. This validation could include clinical factors, quality and safety data, and other ways of demonstrating the validity of certain measures (Brook, McGlynn, & Shekelle, 2000). PRIDIT applied to process measures, shows quality in terms of hospital performance of these process measures, as shown in Lieberthal, 2008. A small number of hospital characteristic variables also have a proven relationship with quality; PRIDIT is able to demonstrate how well these variables proxy for quality in terms of process. Process measures also serve as a measurable proxy for outcomes. With PRIDIT aggregating these process measures, a proxy for high quality outcomes can also be generated. However, it is important to consider alternative implications for these measures, as PRIDIT is a proxy measure and does not determine causation. For example, according to the “teaching to the test” phenomenon, hospitals focus their efforts on what they are being measured on (Werner & Bradlow, 2006), instead of focusing on other areas that may need improvement. Processes that align with high quality care may reflect on a certain patient population, rather than performance.

One other important aspect of the use of PRIDIT for process measures is that these measures come for four categories: heart attack, heart failure, pneumonia, and surgical care improvement. The three conditions, and the improvement of surgical care, are important to a wide variety of patients (U.S. Department of Health and Human Services, n.d.). However, the categories are by no means inclusive of all aspects of hospital performance. The modern general hospital performs a wide variety of procedures across a large number of diseases, so measuring the entirety of hospital performance is beyond the scope of this study. Therefore, these measures are important, but still a limited subset of all possible measures.

HCAHPS data are a way of gauging patient satisfaction. As a result, PRIDIT when applied to this data returns the quality of hospitals viewed through the prism of patient satisfaction. There is a heated debate as to whether patient satisfaction is meaningful, or even negatively correlated with quality. For example, in the case of distance to hospitals, patients will often choose a closer hospital rather than travel for higher quality care (Geweke, Gowrisankaran, & Town, 2003).

Outcomes measures reflect the results of process and structural measures; in this case, we focus on mortality and readmission rates. The outcomes measures are comprised of rates and patient counts. Rates reflect the raw measure, such as mortality rate in a hospital, as well as risk adjustment, including adjustment for the fact that cases that are more difficult are more likely to result in mortality. This data is risk adjusted so that hospitals with more difficult cases are not penalized on their outcomes scores. Patient counts reflect the total number of patients seen in the hospital with a particular condition, such as heart failure, which meet the inclusion criteria for the Hospital Compare database. PRIDIT then returns the quality of hospitals viewed through the prism of risk adjusted hospital outcomes when it is applied to risk adjusted data.

We chose to include both rates and patient counts when applying PRIDIT to outcomes data. It is possible to model only the rates, only the counts, or both. Rates show the outcomes patients care about, while patient counts are a measure of volume. The proven volume-outcome relationship suggests that higher volume hospitals have better outcomes all else being equal. A second reason is the amount of variation in the data. There is not a large amount of variation in the rate data because of the risk adjustment process. This fact has led Medicare to find that most hospitals are “not different from the U.S. average"(Silber et al., 2010) On the other hand, there is a wide amount of variation in the count data, including counts for some hospitals on some conditions of over 2000.

D. Sign choice problem

The sign choice problem in PRIDIT refers to the fact that PRIDIT is a relative, not an absolute, measure of the latent factor of quality. To take the example in Table 10, it would be mathematically equivalent for the scores to all be the opposite sign. The average would have remained at zero, and the results would have been the same but for interpretation—negative scores would have meant higher quality. To put it another way, PRIDIT is sign invariant due to the nature of principal component analysis, where there are two equivalent solutions, one the negative of the other, to any eigensystem.

The reason for this problem is that there is no "fixed point" in the interpretation of hospital quality, i.e., no $Q^{\dagger}$. For example, in the related case of fraud detection, there are certain cases that are adjudicated as being fraudulent that can be used to validate the classification of claims. There is no such equivalent in the quality literature. As a result, it is necessary to select certain quality measures, and decide what their sign is. Then, all other signs are determined by that choice. We chose mortality rates. In any PRIDIT results where mortality rates are included, those variables should always appear as negative: higher mortality rates are worse. That sets an objective standard, as well as a good sanity check, since all variable weights for mortality should be the same, negative, sign. It also points out the difficulties in the quality literature addressed by PRIDIT. Combining many variables of different types allows us to contribute to the debate on which variables are indicators of, and cause, higher quality. Interpretation may still be debated even with quantified results such as the ones we produce.

IX. Appendix 2—additional results

A. Hospital Compare data

Type	Acute care	Critical access	Total
Government	765	391	1,156
Proprietary	753	53	806
Non-profit	2,083	610	2,693
Grand total	3,601	1,054	4,655

Table 16: Count by acute/critical access

Type	Acute care	Critical access	Total
Government			
General	636	391	1,027
VA	129	0	129
Total government	765	391	1,156
Proprietary	753	53	806
Non-profit	2,083	610	2,693
Grand total	3,601	1,054	4,655

Table 17: Count by acute/critical access with VA separate

Type	Acute care		Critical access		Total	
	ER services	No ER services	ER services	No ER services	ER services	No ER services
Government						
General	616	20	386	5	1,002	25
VA	129	0	0	0	129	0
Total government	745	20	386	5	1,131	25
Proprietary	659	94	53	0	712	94
Non-profit	2,015	68	600	10	2,615	78
Total	3,419	182	1,039	15	4,458	197

Table 18: Count by acute/critical access and by ER

Measure code	Not reporting	Reporting	Average
AMI_1	967	3,688	96
AMI_5	1,047	3,608	94
AMI_2	1,059	3,596	95
AMI_3	1,653	3,002	94
AMI_4	1,845	2,810	97
AMI_8a	3,035	1,620	87
AMI_7a	4,208	447	48

Table 19: AMI process measures (sorted by number reporting)

Measure code	Not reporting	Reporting	Average
HF_2	400	4,255	92
HF_1	416	4,239	84
HF_3	560	4,095	91
HF_4	683	3,972	94
PN_2	330	4,325	90
PN_6	354	4,301	91
PN_5c	364	4,291	94
PN_4	401	4,254	93
PN_7	429	4,226	88
PN_3b	454	4,201	94

Table 20: HF and PN process measures (sorted by number reporting)

Measure code	Not reporting	Reporting	Average
SCIP_INF_6	857	3,798	99
SCIP_INF_1	894	3,761	95
SCIP_INF_2	896	3,759	96
SCIP_INF_3	901	3,754	94
SCIP_VTE_1	919	3,736	91
SCIP_VTE_2	925	3,730	90
SCIP_INF_9	954	3,701	88
SCIP_CARD_2	1,167	3,488	91
SCIP_INF_4	3,420	1,235	93

Table 21: SCIP process measures (sorted by number reporting)

Condition	Not reporting	Reporting	Average
PN readmission	412	4,243	18
PN death	430	4,225	12
HF readmission	591	4,064	25
HF death	659	3,996	11
HA death	1,827	2,828	16
HA readmission	2,211	2,444	20

Table 22: Mortality and readmissions rates per hundred patients (sorted by number reporting)

Condition	Not reporting	Reporting	Average
PN readmission	204	4,451	239
PN death	204	4,451	248
HF readmission	217	4,438	292
HF death	223	4,432	242
HA death	339	4,316	125
HA readmission	402	4,253	125

Table 23: Mortality and readmissions volume (sorted by number reporting)

Measure	Response	Not reporting[5]	Reporting	Average
Clean room	Always	828	3,827	71
	Usually	828	3,827	19
	Sometimes / Never	828	3,827	9
Area around patients' rooms kept quiet at night	Always	828	3,827	58
	Usually	828	3,827	30
	Sometimes / Never	828	3,827	11
Doctors communicate	Always	828	3,827	80
	Usually	828	3,827	15
	Sometimes / Never	828	3,827	5
Nurses communicate	Always	828	3,827	76
	Usually	828	3,827	19
	Sometimes / Never	828	3,827	5
Patients receive help quickly	Always	828	3,827	64
	Usually	828	3,827	25
	Sometimes / Never	828	3,827	11
Patients' pain well controlled	Always	828	3,827	69
	Usually	828	3,827	23
	Sometimes / Never	828	3,827	7

[5] Includes 129 VA hospitals with HCAHPS data not included in Hospital Compare.

Measure	Response	Not reporting[5]	Reporting	Average
Patients given information about what to do during their recovery	Yes	830	3,825	82
	No	830	3,825	18
Staff explain about medicines	Always	832	3,823	61
	Usually	832	3,823	18
	Sometimes / Never	832	3,823	21
Patients rate the hospital overall	9-10	829	3,826	68
	7-8	829	3,826	23
	0-6	829	3,826	9
Would patients recommend the hospital to friends and family?	Yes, definitely	829	3,826	70
	Yes, probably	829	3,826	25
	No	829	3,826	5

Table 24: HCAHPS measures

Collinear HCAHPS measures:

- Patients overall rating of the hospital (three levels, two variables)
- Would patients recommend the hospital to friends and family? (three levels, two variables)
- Staff explain about medicines (three levels, two variables)
- Patients given information about what to do during their recovery (two levels, one variable)
- Patients' pain well controlled (three levels, two variables)
- Patients receive help quickly (three levels, two variables)
- Doctors communicate (three levels, two variables)
- Nurses communicate (three levels, two variables)
- Clean room (three levels, two variables)
- Area around patients' rooms kept quiet at night (three levels, two variables)

B. AHA data

Type	Count
No affiliation	3,424
COTH only	8
Medical school only	313
Residency only	5
COTH and medical school	11
COTH and residency	3
Residency and medical school	581
COTH, residency and medical school	310
Total	4,655

Table 25: Teaching status counts

Type	Count
No accreditation	1,356
DNV	114
Joint Commission	3,159
Joint Commission and DNV	26
Total	4,655

Table 26: Accreditation counts

Type	Count
Community hospital	4,394
Network	1,414
Cluster	2,586
Beds (average)	172

Table 27: Other counts

C. PRIDIT results—all data

Measure type	Measure	Weighting	Rank
Demography	ER	0.24	33
Demography	Acute Care	0.43	25
Demography	VA	0.05	58
Demography	NPF	0.16	37
Demography	FP	−0.06	53
Demography	Community	0.05	57
Demography	Network	0.10	44
Demography	Cluster	0.19	34
Demography	JC Accreditation	0.40	27
Demography	ACGME	0.40	28
Demography	Med School	0.38	29
Demography	COTH	0.28	31
Demography	DNV Accreditation	0.02	60
Demography	Beds	0.72	12
Process	AMI_1	−0.07	52
Process	AMI_2	−0.05	56
Process	AMI_3	−0.09	51
Process	AMI_4	0.01	69
Process	AMI_5	−0.05	55
Process	AMI_7a	0.01	65
Process	AMI_8a	−0.02	61
Process	HF_1	0.11	43
Process	HF_2	0.27	32
Process	HF_3	0.01	64
Process	HF_4	0.09	49
Process	PN_2	0.10	46
Process	PN_3b	0.00	70
Process	PN_4	0.13	38

Measure type	Measure	Weighting	Rank
Process	PN_5c	−0.12	41
Process	PN_6	0.09	48
Process	PN_7	−0.01	68
Process	SCIP_CARD_2	0.01	67
Process	SCIP_INF_1	0.04	59
Process	SCIP_INF_2	−0.10	47
Process	SCIP_INF_3	−0.12	39
Process	SCIP_INF_4	−0.01	66
Process	SCIP_INF_6	0.00	71
Process	SCIP_INF_9	−0.12	40
Process	SCIP_VTE_1	0.02	63
Process	SCIP_VTE_2	−0.02	62
Outcome	HA Mortality Rate	−0.12	42
Outcome	HA Mortality N	0.71	13
Outcome	HA Readmission Rate	0.06	54
Outcome	HA Readmission N	0.70	16
Outcome	HF Mortality Rate	−0.17	35
Outcome	HF Mortality N	0.71	14
Outcome	HF Readmission Rate	0.10	45
Outcome	HF Readmission N	0.71	15
Outcome	PN Mortality Rate	−0.09	50
Outcome	PN Mortality N	0.65	18
Outcome	PN Readmission Rate	0.17	36
Outcome	PN Readmission N	0.65	19
Satisfaction	Always clean	−0.73	9
Satisfaction	Usually clean	0.73	11
Satisfaction	Nurses always communicated well	−0.80	2
Satisfaction	Nurses usually communicated well	0.77	4

Measure type	Measure	Weighting	Rank
Satisfaction	Doctors always communicated well	−0.79	3
Satisfaction	Doctors usually communicated well	0.75	7
Satisfaction	Patients always received help	−0.84	1
Satisfaction	Patients usually received help	0.76	6
Satisfaction	Pain was always well controlled	−0.74	8
Satisfaction	Pain was usually well controlled	0.61	21
Satisfaction	Staff always explained	−0.77	5
Satisfaction	Staff usually explained	0.36	30
Satisfaction	Staff gave recovery information	−0.41	26
Satisfaction	Rated 7-8	0.62	20
Satisfaction	Rated 9-10	−0.68	17
Satisfaction	Always quiet	−0.73	10
Satisfaction	Usually quiet	0.54	22
Satisfaction	Definitely recommend	−0.50	23
Satisfaction	Probably recommend	0.44	24

Table 28: PRIDIT ranked variables

D. PRIDIT results—individual measure types

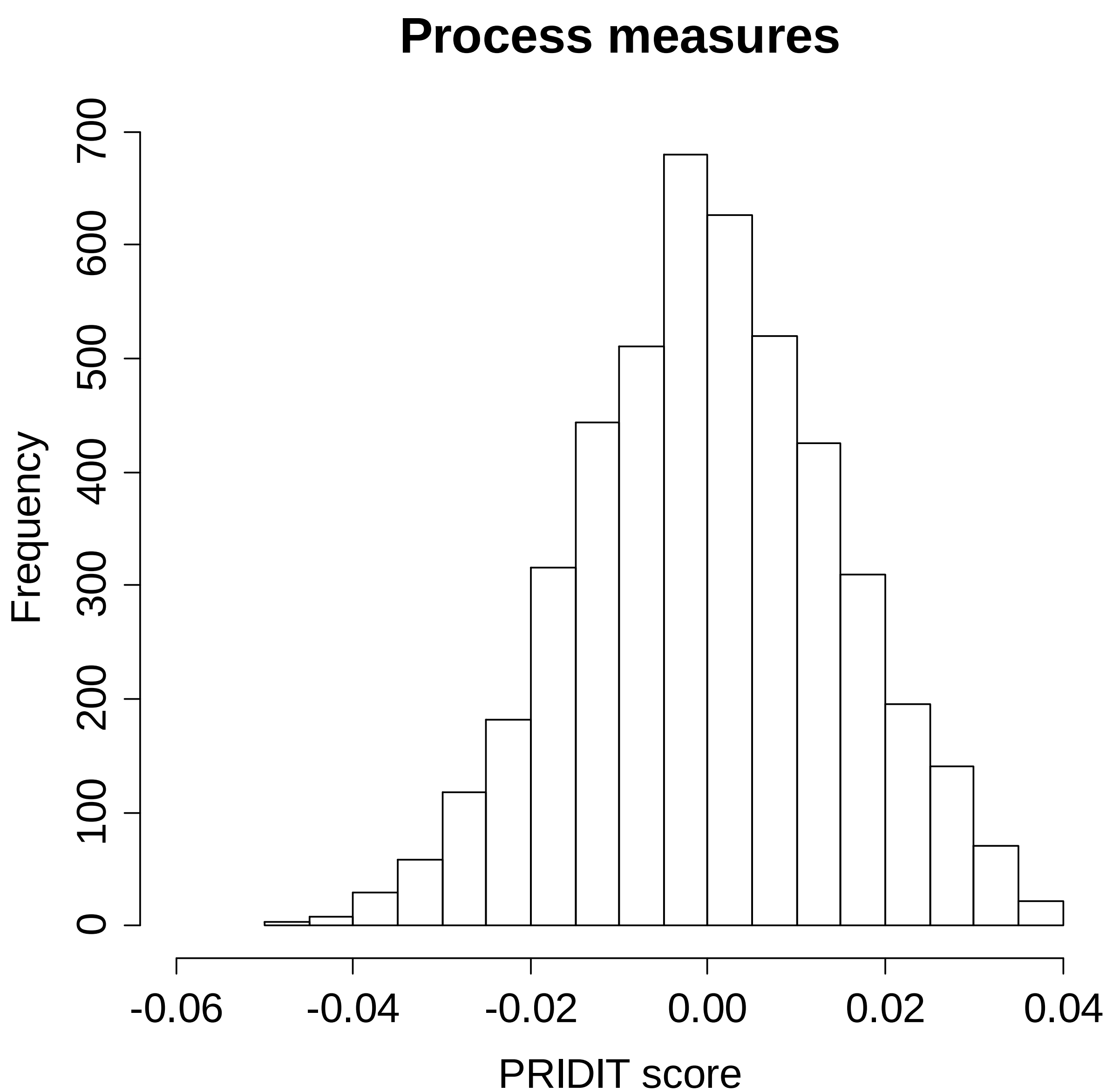


Figure 7: Process measure PRIDIT score distribution

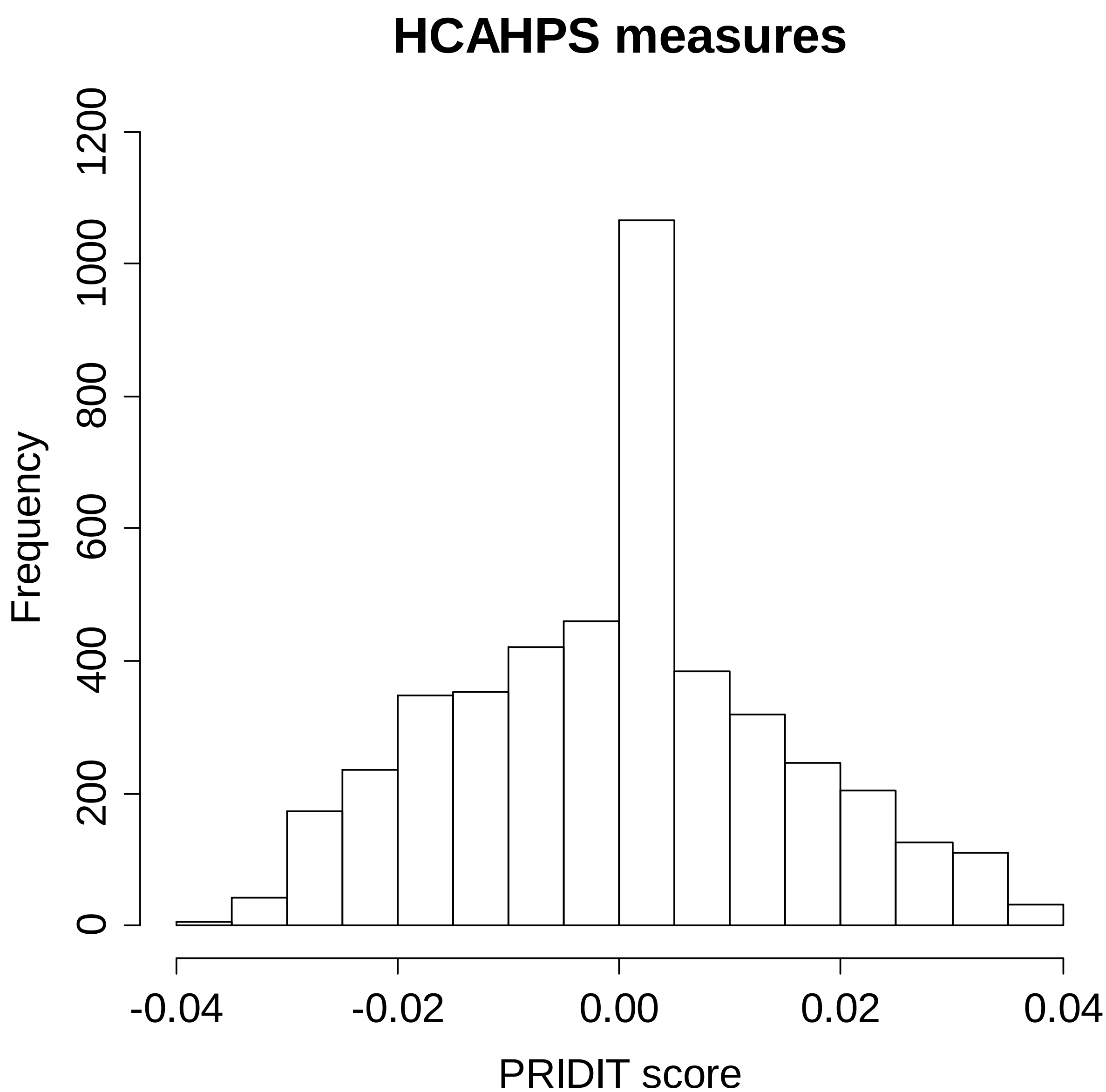


Figure 8: HCAHPS measure PRIDIT score distribution

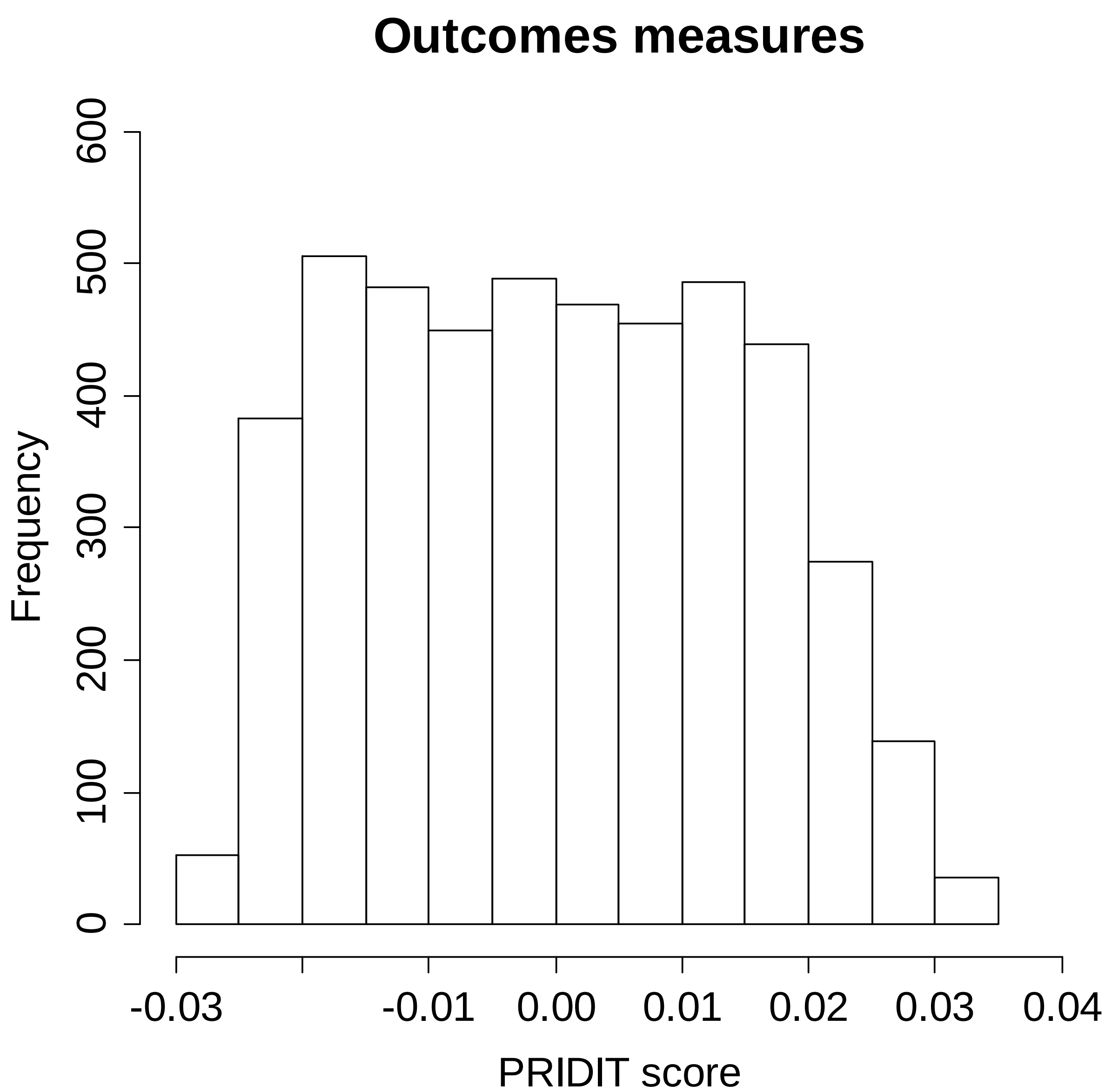


Figure 9: Outcomes measure PRIDIT score distribution