

2013

What are the characteristics that explain hospital quality? A longitudinal Pridit approach

Robert D. Lieberthal
Thomas Jefferson University

Dominique M. Comer
Thomas Jefferson University

Follow this and additional works at: <https://jdc.jefferson.edu/healthpolicyfaculty>

 Part of the [Health and Medical Administration Commons](#)

[Let us know how access to this document benefits you](#)

Recommended Citation

Lieberthal, Robert D. and Comer, Dominique M., "What are the characteristics that explain hospital quality? A longitudinal Pridit approach" (2013). *College of Population Health Faculty Papers*. Paper 73.

<https://jdc.jefferson.edu/healthpolicyfaculty/73>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in College of Population Health Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

As submitted to:

Risk Management and Insurance Review

And later published as:

What are the characteristics that explain hospital quality? A longitudinal Pridit approach

Volume 17, Issue 1, pages: 17-35

2013

DOI: 10.1111/rmir.12017

Robert D. Lieberthal^{1, 2}
Assistant Professor
Thomas Jefferson University

Dominique M. Comer¹
Health Economics and Outcomes Research Fellow
Thomas Jefferson University

April 5, 2016

Health outcomes vary substantially between high and low quality institutions, meaning the difference between life and death in some cases. Prior literature has identified a number of variables that can be used to determine hospital quality, but methodologies for combining variables into an overall measure of hospital quality are not well developed. This analysis builds on the prior investigation of hospital quality by evaluating a method originally developed for the detection of healthcare fraud, Pridit, in the context of determining hospital quality. We developed a theoretical model to justify the application of Pridit to the hospital quality setting and then applied the Pridit method to a national, multiyear dataset on U.S. hospital quality variables and outcomes. The results demonstrate how the Pridit method can be used predictively, in order to predict future health outcomes based on currently available quality measures. These results inform the use of Pridit, and other unsupervised learning methods, in fraud detection and other settings where valid and reliable outcomes variables are difficult to obtain. The empirical results obtained in this study may also be of use to health insurers and policymakers who aim to improve quality in the hospital setting.

¹ Thomas Jefferson University, Jefferson School of Population Health, 901 Walnut St, 10th Floor, Philadelphia, PA 19107. Rob Lieberthal can be reached at robert.lieberthal@jefferson.edu. Dominique Comer can be reached at dominique.comer@jefferson.edu

² Please direct all correspondence to Robert Lieberthal, Thomas Jefferson University, Jefferson School of Population Health, 901 Walnut St, 10th Floor, Philadelphia, PA 19107. Phone: (215) 503-3852, Fax (215) 923-7583, email: robert.lieberthal@jefferson.edu

Keywords: Hospital quality, Pridit, predictive modeling, unsupervised learning

Acknowledgements: The Society of Actuaries provided funding for this work through their Health Section. Part of Dr. Comer's time spent on this research was funded by a Postdoctoral Fellowship award on Health Outcomes from the PhRMA Foundation. We would like to thank the Society of Actuaries Project Oversight Group (POG), which provided advice and guidance during the course of the project. Katie O'Connell also provided valuable research assistance. A version of this work was originally published as a report by the Society of Actuaries under the title "Validating the PRIDIT Method for Determining Hospital Quality with Outcomes Data" and presented at the 2013 ARIA Annual Meeting. Muhammed Altuntas provided valuable comments and feedback.

I. Introduction

A. Background

Hospitals are a critical setting for healthcare quality improvement in the U.S.; 31% (\$814 billion) of the \$2.6 trillion of healthcare delivered in 2010 was spent in the hospital (Martin, Lassman, Washington, & Caitlin, 2012). Quality of care is quite variable throughout the U.S., with much variation in services provided. The overuse and underuse of such services has been identified as a critical problem within the U.S. healthcare system (Agency for Healthcare Research and Quality (AHRQ), 2002). Medical errors in the hospital setting that may result from poor quality of care account for approximately \$17 billion each year (Van Den Bos et al, 2011). Thus, any methodology that can provide evidence about the overall quality of hospitals, their trends in quality over time, and the variables that indicate high quality care have the potential to improve the quality and lower the costs of U.S. healthcare.

One major challenge in the study of overall hospital quality is that general hospitals provide a wide variety of services and perform a number of different functions therein. Hospitals care for patients with a range of chronic and acute conditions. Further, adding in the complexity of the U.S. healthcare system, investigating the quality and the outcomes of healthcare has become substantially more difficult.

Health insurers have historically played a limited part in the push to improve quality, but their efforts are growing. An example of this is the growing aspect of Pay for performance programs as a part of many managed care contracts. A number of quality improvement efforts are also rapidly appearing at the national level, such as the National Quality Forum's listing of "Never Events" –medical errors that should never occur—and policy recommendations to stop paying for these largely preventable occurrences. Medicare has used risk-sharing arrangements to redistribute withheld money from hospitals to those that meet certain benchmarks in mortality and readmissions rates. Despite these changes, payers still frequently pay for care that is substandard; it is a common industry practice to reimburse hospitals for corrective care, which reduces the incentive for hospitals to increase their quality of care. Only recently have insurers begun to define specific "Never Events" that they will not pay for (Milstein, 2009)

B. Literature review

When the hospital is the unit of observation, determining “high quality” is a challenge. Objectifying the quality of hospital has proved to be a difficult and controversial topic. Multiple methodologies exist for creating measures for hospital processes and outcomes (i.e. Lovaglio, 2012; Shahian, Wolf, Iezzoni, Kirle, & Normand, 2010). Despite this disagreement in measuring quality, multiple programs and interventions exist that attempt to improve hospital quality. Programs such as Pay for Performance and Meaningful Use utilize financial incentives and disincentives 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 direct comparisons between 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 overall quality. In other words, while many analyses focus on quality by hospital type, on improving the processes of care delivery, or on improving healthcare outcomes, few prior studies have combined these types of analyses into an overall picture of hospital quality.

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 of care data to measure quality. The Pridit method is well suited to accomplishing this prioritization process of quality measures. Pridit is also able to utilize many types of variables, some of which may not be as useful in determining quality (e.g. parking costs, food quality, visiting hours, etc.), some of which may be useful proxies (e.g. how often aspirin is administered after a heart attack when indicated), and some that patients and other stakeholders 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.

Pridit can also be considered as one method within the larger set of methodologies known as cluster analysis. Derrig (2002) develops a “claim sorting algorithm development flow” that includes various methods of cluster analysis in Step 4 including Kohonen’s self-organizing feature map, Pridit, and fuzzy methodologies. This analysis applies the highest level of claim sorting proposed by Derrig (Step 8, Dynamic Testing) by applying Pridit at data observed at multiple times. Additional applications of cluster analysis in the insurance context include cluster analysis to compare different insurers, such as Berry-Stölzle & Altuntas (2010). The Pridit methodology shares common features with this prior use of cluster

analysis, as it essentially “...standardize(s) each variable by subtracting its mean...” The major difference with the Pridit method is that instead of dividing by a variable’s own standard deviation, Pridit uses Principal Components Analysis to standardize each variable by its standard deviation as well as its correlation with all other variables in the data set, as represented by the first eigenvector associated with the PCA system. We describe this in detail in section II.B below.

C. Motivation

Presently, there are large, longitudinal hospital quality data sets that were not available even five years ago. This availability allows for the validation of Pridit scores using a variety of data sources against one another. Specifically, we can generate a rich set of hospital scores using demographic, process, and patient satisfaction data, and compare the results to outcomes measures. We can then compare scores over time to judge the stability and predictive nature of Pridit.

Given the data available, our motivation in exploring the application of Pridit to hospital quality in this study was to expand on previous analyses by including multiple types of variables used to score hospital quality. Our goal was to determine whether the aggregation of many different types of quality data led to the generation of stable quality scores over time. Thus, our more general aim was to explore the validation of Pridit in the hospital quality. A stable scoring system can facilitate efforts by health insurers to implement pay for performance programs and risk sharing arrangements.

One difficulty with the use of unsupervised learning methods, which is especially acute in the field of fraud detection, is that often there are no standard outcomes measures available. Fraudulent cases are settled quietly and data is highly proprietary, possibly restricted to use by a small subset of insurance company employees. Our use of outcomes variables as part of the Pridit analysis allows us to draw conclusions about the use of Pridit in the hospital settings, where data on inputs and outcomes is publically available. Thus, a secondary motivation of this analysis was to draw conclusions about the use of the Pridit method in a setting where outcomes measures are available (hospitals), in order to draw inferences about its performance in settings where outcomes measures may be difficult to obtain (insurance fraud).

D. Theoretical model³

In our theoretical model, we suppose that there is an unobserved, latent measure of relative hospital quality, Q . This measure is ordinal and 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 . In other words, for any variable that represents a measure of high quality healthcare, hospital i is likely to score higher than hospital j , all else equal.

We conceptualize the random variable q as one such measure of hospital quality. q is a real valued scalar that represents the quality of some aspect of hospital care. q may also have a more limited number of possible values, such as being a binary or categorical variable. The fact that q is a random variable reflects the fact that quality is higher in a probabilistic sense: the distribution of q at the higher quality hospital i has a larger mean value than at the lower quality hospital j .

Now, we compose a vector of quality measures \mathbf{q} with n elements. Each member of \mathbf{q} , q_1, q_2, \dots, q_n , is a scalar proxy for quality. That is, the correlation of the k^{th} measure of quality with overall quality, $\text{corr}(q_k, Q)$ is on 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 , though there may be some 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 \mathbf{q} where the proxy is a better measure of quality than any observed proxy, i.e. $\text{corr}(Q^*, Q) > \text{corr}(Q^\dagger, Q)$, all else being equal. Note also that if Q is real valued, then it is possible to rescale Q onto the interval $(-1, 1)$ without loss of information.

Given this setup, Brockett et. al., 2002, developed Pridit, a methodology that produces Q_j^* , a single number that represents the latent variable Q_j . 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 score is normed to 0 , so that negative scores represent membership in the “suspicious” class (meaning low quality in the hospital context), and positive scores are in the “non-suspicious” class (meaning high quality in the hospital context). The scale is also multiplicative; a score of 0.50 is twice as strong, in terms of indicating the latent factor, as that with a score of 0.25 . On an absolute value basis, the scale is

³ This section is based largely on Appendix 1 of Lieberthal & Comer, 2013.

also multiplicative with negative values. A positive 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. 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 (Brockett, Derrig, Golden, Levine, & Alpert, 2002; Ai, Brockett, Golden, & Guillen, 2012).

II. Methodology

A. Data

The main source of data collected for this study came from the Hospital Compare database, available for download via the Centers for Medicare and Medicaid Services’ (CMS) Medicare.gov website (Centers for Medicare and Medicaid Services’ (CMS), 2012).⁴ Medicare claims and enrollment data comprise the majority of this database. Hospital Compare focuses on three disease states in their reporting: heart attack, heart failure, and pneumonia. Additional demographic data from the American Hospital Association supplemented our overall data set (American Hospital Association (AHA), 2011). The variables in our data set can be categorized in four groups: demographic, process, outcomes, and patient satisfaction.

Demographic measures show the characteristics of hospitals, which are properties that generally remain fixed across time. These measures include ownership status (for profit, nonprofit, government) and teaching status. Teaching status represents a number of activities that hospitals may engage in to train new physicians. Teaching status also has financial implications for hospitals; these hospitals may receive additional financing to cover the cost of teaching above and beyond the cost of patient care. Demographic measures in our data also consist of accreditation status. The majority of U.S. hospitals are accredited, which requires an upfront and ongoing cost to the hospital, and generally results in higher reimbursement for the hospital. Other hospital characteristics include membership in a hospital network, being part of a cluster of hospitals, and number of beds. These measures cover the size and scope of hospitals, both in terms of how large they are, and whether they are part of a larger health system. Our data includes 14 binary and continuous variables that measure demographic characteristics of hospitals.

⁴<http://www.medicare.gov/hospitalcompare>

Process measures capture the actions performed within the hospital and reflect the care that the hospital provides to patients. In the health care system, these measures can represent actions such as smoking cessation counseling and the timing of appropriate antibiotics. Hospital Compare collects process measures in the following areas of health care: heart attack, heart failure, pneumonia, and quality of surgical care as measured through the Surgical Care Improvement Project (SCIP). In total, our data includes 26 process measures. Each takes an integer value from 0 to 100 representing between 0% and 100% adherence to a particular process measure. Hospital Compare will only report process measures when there are at least 25 cases to base a measure on; in other cases the variable value is empty (“N/A”).

Outcomes measures capture the results of care given to patients. In contrast to the 12 month collection period for process and patient satisfaction measures, the collection period for outcomes measures is 36 months. The mortality and readmission rates are reported as 30-day risk-adjusted rates, with a continuous value from 0 to 1. Hospital Compare will only report outcome measures when there are at least 25 cases to base a measure on; in other cases the variable value is empty (“N/A”). Hospitals also report a patient count for each measure. Hospital Compare reports all patient counts regardless of whether the number of cases is above or below 25. Our data includes 12 outcome and volume measures⁵.

Patient satisfaction measures in Hospital Compare were obtained through Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). HCAHPS is a standardized survey instrument used to measure patients’ perspectives in hospital care. The questions asked in the survey span a variety of patient experiences. Satisfaction is a form of patient reported outcome, in which the source of the measure comes directly from the patient’s perspective. Hospital Compare will only report satisfaction measures when there are at least 25 cases to base a measure on; in other cases the variable value is empty (“N/A”). The survey contains ten questions, each of which has a rank ordered response from patients representing better or worse experiences. Patient responses are then collapsed by Hospital Compare into three categories—low (0-6), medium (7-8), and high (9-10). Hospital Compare reports an integer value between 0 and 100 representing the percent of patients whose response falls into a

⁵ Volume measures could be considered either a demographic measure or an outcome measure. Larger hospitals will tend to have higher volumes, and volumes will also tend to vary with the ebb and flow of patients into a particular hospital during a particular time period. The question of how useful volume is as an indicator of quality is an open question in the literature.

specific category for a given question. We chose one reference response for each question and excluded it as perfectly collinear with the other measures for the same question, ultimately using 19 of the 29 HCAHPS measures. Of note, Veterans' Affairs hospitals do not report HCAHPS measures to Hospital Compare, and thus their satisfaction scores are not included in the analysis.

We combined the Hospital Compare and American Hospital Association data to create our study dataset for 2011. We similarly also created datasets for the years 2010, 2009, and 2008, using Hospital Compare data from each year respectively in combination with the American Hospital Association demographic data from 2011. In some instances, 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 as that was the year Hospital Compare began to collect outcomes measures⁶.

B. Analysis

In our analysis plan, we applied the Pridit model as detailed in the theoretical model to our dataset. Our first step was to apply the Pridit model to the full dataset for 2011. We generated a single score between -1 and 1 showing hospitals' relative quality. We considered this score to be a ranking of the relative success of hospitals, with more positive numbers measuring higher performance. The composite score measures the variance of variables individually and their covariance with other measures using the Principal Components Analysis (PCA). Pridit selects the first component that explains the greatest degree of variation observed in the data. When we applied Pridit to the 2011 data, we also generated a score between -1 and 1 showing each variable's relative weight in determining the overall score. This score is a relative weight reflecting the importance of each variable as determined through PCA, with the sign showing the direction of association between the variable and the overall score. Larger numbers in absolute value terms indicate variables of greater importance. We also generated scores using the 2010, 2009, and 2008 data. Similarly as above, these analyses generated a single score between -1 and 1 showing hospitals' relative performance in each year and a score between -1 and 1 showing the measures' relative weight in determining the score in each year.

We then analyzed the distribution of hospital scores in each year. The mean of Pridit is fixed at zero. The median shows whether the 50th percentile hospital is relatively high quality (positive) or low

⁶ Additional detail is in our final report to the Society of Actuaries, available at <http://www.soa.org/Research/Research-Projects/Health/research-val-pridit-method.aspx>.

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.

In order to assess this performance across time, we calculated the correlation of 2011 scores with 2010, 2009, and 2008 scores. We also calculated 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 results over time generated results as to the stability of hospital performance. We also assessed the stability of the results as a way of testing the validity of Pridit itself for hospital quality. If scores are stable across time, Pridit may be useful as a predictive model of future hospital quality. Thus, assessing correlations of scores and weights over multiple years is a test of the validity of our model in the setting of hospital quality.

III. Results

A. Cross sectional results

Our first set of results show the Pridit scores generated for 2011. The histogram for overall results in Figure 1 shows the overall distribution of hospital scores. 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 $(-0.015, 0.015)$. The standard deviation as measured in the data was approximately 0.01 as shown in Table 1. Most hospitals were of average quality and the median hospital was just below average quality. Consistent with prior results is our finding that the median hospital quality score is below the average score of zero. The tendency of hospitals to be below average is also reflected in the negative skewness we found in the scores. Being able to better examine these slightly below average hospitals shows the usefulness of utilizing a large measure set for Pridit. For example, we identified certain qualitative differences between average, high quality, and low quality hospitals. Smaller, independent, non-teaching hospitals tended to be on the lower end of the range of scores. We also found that the distribution of hospitals' scores included a large number of hospitals in the "low quality" class whose scores were negative, but very close to zero. In other words, the membership of many hospitals in this class is weak based on the

data. This was also demonstrated by the standard deviation in the data. The fact that the range where most low quality hospitals are found, $(-0.015,0)$ is less than two standard deviations wide shows that Pridit was not able to distinguish these hospitals with a great degree of precision at a certain point in time.

The full variable ranking can be found in Table 2. We highlighted the top ten measures in terms of the absolute value of their weights. One significant finding was that all of the top ten measures of quality are consumer satisfaction scores from HCAHPS measures. The highest-level consumer satisfaction scores were negatively associated with quality, while middle level scores were positively associated with quality (the lowest level was the reference category). In many cases, the variable weight for the middle level response was of similar magnitude, but was in the opposite direction of the corresponding high-level response. Taken together, these variable weights largely cancel out. Thus, the contribution of patient satisfaction variables to scores is less than the individual variable ranks imply.

In looking at the outcomes measures, the measure weightings showed that while mortality rates were negatively associated with quality, readmission rates were positively associated with quality. The weights on both of these measures are relatively small when compared to total patient counts eligible for measurement on a particular outcome (mortality or readmission). Thus, when assessing hospital quality, risk adjusted outcomes are less informative than volumes. The characteristics variable “Number of beds” showed the same important relationship, where larger hospitals had higher quality in addition to the outcome count variables.

B. Longitudinal results

The values of scores using all datasets from past years were highly correlated across time. The correlation coefficient for the 2010 and 2011 scores was 0.93. The scores in a given year were highly predictive of future performance in terms of the score. In our comparison of dispersion of scores using the 2010 and 2011 data, there was also a high degree of consistency. There were a large number of slightly below average hospitals in both years. There were also small numbers of extremely high and extremely low quality hospitals. There was a bimodal distribution in both years; however, in 2011, the large mass of hospitals was farther from average (lower quality) than in 2010. Thus, the lower quality hospitals are easier to distinguish from the average in 2011 than in 2010 (see Figure 2 and Figure 3). It should be noted that 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 of hospitals not available for 2010 was small (98), this fact is likely a minor driver of the results.

In addition to the hospitals' weights, the measures' weights were highly correlated over time (correlation coefficient > 0.99), demonstrating again the stability of scores over time. Examining Figure 4 shows the types of variable weightings for determining quality both at a point in time and across multiple years. We note first that there are many variables clustered around zero. Pridit positively weighted many variables for higher quality, and weighted fewer variables negatively for poorer quality. That is consistent with the fact that most process of care and patient satisfaction variables were designed to be positively associated with quality.

Next, we examined the correlation of outcomes measures to all scores derived from the dataset. We examined the correlation between 2010 scores using all variables and 2011 outcomes measures to determine whether Pridit predicted outcomes in addition to future scores. The correlations of quality scores and heart attack, heart failure, and pneumonia mortality rates were -0.19, -0.20, and -0.11, respectively, as shown in Table 3. 2010 scores were more highly correlated with 2011 outcomes than the 2011 scores. These correlations reflect the intent of the use of mortality outcomes measures—higher quality hospitals should have lower mortality rates. The correlations with heart attack, heart failure, and pneumonia readmissions were 0.12, 0.10, and 0.17, respectively. These correlations do not reflect the intent of the use of readmissions rates—higher quality hospitals are thought to find ways to have lower than average readmissions rates. We found similar degree of correlation between 2009 scores and 2011 outcomes, and smaller correlations between 2008 scores and 2011 outcomes. Based on these results, it took about three years for the predictive power of Pridit to decline significantly.

We also demonstrated the difference between using the full measure set of all variables and partial measure sets utilizing only certain types of variables. The use of only process and demographic measures showed a consistent but less highly correlated view of outcomes. The correlations of demographic and process variable based scores with heart attack, heart failure, and pneumonia mortality were all -0.09. The correlations of demographic and process variable based 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 did not increase the correlation, showing that we could judge hospitals on process measures alone for this disease state. For readmissions, process measures seemed to have no bearing on risk adjusted readmissions rates.

The use of only patient satisfaction and demographic variables showed very different results. The correlations of demographic and HCAHPS scores with heart attack, heart failure, and pneumonia mortality were 0.08, 0.15, and 0.04, respectively. Higher satisfaction was associated with higher mortality rates. The correlations of demographic and HCAPS scores with heart attack, heart failure, and pneumonia readmissions were -0.12 , -0.13 , and -0.16 , respectively. Higher satisfaction was strongly correlated with a lower likelihood of risk-adjusted readmission. These hospitals also have much lower volumes, with correlations with the patient count measures on the range $(-0.34, -0.30)$. High satisfaction hospitals have lower volumes, lower readmissions, and worse mortality. The results are broadly similar when we added process measures of care, showing that satisfaction variables dominate process variables in calculating scores.

IV. Discussion

A. Exploring the validation of Pridit

The use of multiple measures of quality over time adds significantly to the point in time estimates of Pridit scores. The high degree of consistency in scores and serial correlation of scores over time allowed us to better characterize the quality of hospitals. We demonstrated that scores are more accurate than they appear from the point of time estimate. As a random draw from the distribution in the 2011 column of Table 1, many hospitals cannot be precisely placed into the low quality or high quality class.

The distribution of hospitals is similar across the years 2008-2011. This, combined with the high degree of correlation of scores as shown in Figure 5, demonstrated that the results based on the full set of variables in one year are likely valid in the next year. The intention of Pridit application when using all applicable data is to give an overall picture of hospital quality. This overall picture will include all of the elements that are input into it—demographic, process, outcome, and satisfaction measures. How well hospitals score on the variables they report will determine the quality of the hospital, especially on 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 many measure types will tend to get the highest weights. Similarly, demographic characteristics, such as not-for-profit status, will also affect multiple types of performance. Thus, it is possible for the Pridit score to give a broad a view of hospital performance.

B. Implications for health insurance

One major finding of this study is that patient satisfaction is a poor measure of quality. The implications of our satisfaction results when combined with other measures of quality were twofold. First, the best hospitals were not the ones that were the quietest or that had the most responsive clinicians. Busier hospitals tended to have better performance, which is consistent with the volume-outcome relationship (Luft, Hunt, & Maerki, 1987). These hospitals scored highly using process and outcomes variables and indicators of volume, but only in the middle in terms of patient satisfaction. There are two explanations for the pattern of variable weights generated by Pridit. First, Pridit is able to deduce a pattern of correlation by relating high quality to the highest scores on process and outcomes measures and mid-level scores on satisfaction. Secondly, the Pridit method reduces the value of overall patient satisfaction by weighing these two measures in the opposite direction. Pridit utilized the high degree of variation in top achievement in satisfaction and mid-level achievement in satisfaction, and thus ascribed to each a strong, opposite-signed variable weight.

The combination of various types 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. With Pridit, hospitals can focus on collecting the measures that will be most useful in quality improvement efforts. The measures that have the largest impact on quality will also tend to be the most useful measures over time. While there may be a need to replace measures as they become less useful, the process of continually adding more measures to Hospital Compare may not be improving quality. For effective quality care monitoring, the measures that are collected should demonstrate a positive impact on quality care.

As health insurers consider broad strategies for quality improvement and cost control, Pridit may have a role in improving these efforts. Generally, contracting for healthcare is local, where a health insurer may negotiate with a small number of hospitals to provide inpatient healthcare services in a given area. Our results suggest that insurers should consider a wide variety of data and use it to negotiate rates with hospitals, or detect higher quality hospitals for in-network contracting. Insurers should not spend resources, focus, and energy in implementing measure-based pay for performance programs or other more granular hospital performance programs; such programs are better left to the individual hospital.

Our results also suggest that the strategies of reference pricing or centers of excellence may be most useful for insurers to consider. Reference pricing involves a hospital negotiating a maximum rate for a given service with a small number of providers, and then capping its share of costs at that level for all hospitals (Robinson & MacPherson, 2012). Centers of excellence refers to the strategy of selecting a small number of high quality providers, and attempting to drive as much volume to that provider as possible through contracting and other incentives for insured individuals (Robinson & MacPherson, 2012). In both reference pricing and centers of excellence strategies, quality serves as a threshold. In the case of reference pricing, insurers must set a minimum threshold such that the incentives for hospitals is to maximize the value received from expenses for care while maintaining a level of quality. In centers of excellence, payers could select preferred hospitals, conditional on a quality threshold. Pridit is an ideal methodology for both of these applications, as it allows insurers to set any threshold they wish by choosing a minimum Pridit score. This score is not likely to vary a great deal across time or as measures change. In other words, insurers should take quality variation as given, and then follow the implication, which is to pay most hospitals the same amount, to drive patients to those few hospitals that are truly outstanding, and to steer patients away from those few hospitals that are truly of poor quality. Such a strategy may be difficult to implement from the point of view of consumer satisfaction if the hospitals that insured are most satisfied with are not the ones that deliver the best outcomes.

C. Limitations

The use of the results of this analysis is subject to two important limitations. The first is that the data used for this study contains both missing elements and risk adjusted elements. In other words, for certain variables, if there were fewer than 25 encounters, then no value was reported for that variable. We utilized an averaging method for such missing data, assigning them the average value for all hospitals that did report a variable value. Regression based and other method filling in missing data may produce more accurate results, but are beyond the scope of this analysis. For other variables, risk adjusted measures were reported, when the ideal would be to use the raw scores so that we would be able to determine the ideal risk adjustment system to use the variables for Pridit. More broadly, it will always be the case that Pridit is an unsupervised method, so that validation of the results using regression with a defined dependent (left hand side) variable will require the use of additional data and/or other methodologies. We consider the use of such comparisons as an important starting point for additional investigation into the Pridit method.

D. Research implications

Future applications of Pridit include utilizing the methodology presented in this report to analyze other settings for health care. Ideally, similar variables and same methodology would be used to assess quality in the outpatient, pharmacy, and home care setting and to compare which factors and drivers of quality are common across health care settings. In reality, it is difficult to judge different health care settings using the same set of measures; this is reflected in the variety of quality measures that are collected for various sites of care delivery. 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.

In conclusion, Pridit adds to our understanding of hospital quality, and presents as a new methodology insurers can use for contracting, network selection, and pricing. Focusing on the relationship between variables that exist in the data and the construction of a single quality score, Pridit allows us to characterize hospital quality using a rich and diverse dataset. Indeed, our use of multiple outcomes allowed us to show that certain aspects of hospital quality measurement, specifically satisfaction and readmissions, are related to overall hospital performance in a different way than has been traditionally assumed. Thus, analysis that is motivated by questions of how to improve quality overall may be more likely to capture quality improvement than those which are motivated by improving specific measures of quality or certain types of quality variables.

V. Bibliography

Agency for Healthcare Research and Quality. (2002) Improving health care quality: Fact sheet [internet].

Available from <http://www.ahrq.gov/research/findings/factsheets/errors-safety/improving-quality/index.html>

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.

American Hospital Association (AHA). (2011). *AHA Annual Survey Database* [Data file and code book].

Berry-Stölzle, T. R. and Altuntas, M. (2010). A resource-based perspective on business strategies of newly founded subsidiaries: The case of German pensionsfonds. *Risk Management and Insurance Review*, 13(2): 173–193. doi: 10.1111/j.1540-6296.2010.01183.x

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

Centers for Medicare and Medicaid Services (CMS). (2012). *Hospital Compare* [Data file]. Retrieved from <http://www.medicare.gov/hospitalcompare>

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. doi: 10.1177/0272989X10395596

Derrig, R. A. (2002). Insurance fraud. *Journal of Risk and Insurance*, 69(3): 271–287. doi: 10.1111/1539-6975.00026

- 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
- Lieberthal, R. D. and Comer, D. M. (2013). Validating the PRIDIT method for determining hospital quality with outcomes data. Report for the Society of Actuaries. Available from
<http://www.soa.org/Research/Research-Projects/Health/research-val-pridit-method.aspx>
- 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.
- Martin, A. B., Lassman, D., Washington, B., & Catlin, A. (2012). Growth in US health spending remained slow in 2010; health share of gross domestic product was unchanged from 2009. *Health Affairs, 31*(1): 208–219.
- Milstein, A. (2009). Ending extra payment for “never events”—stronger incentives for patients' safety. *New England Journal of Medicine, 360*(23): 2388–2390.
- Robinson, J.C., & MacPherson, K. (2012). Payers test reference pricing and centers of excellence to steer patients to low-price and high-quality providers. *Health Affairs 31*(9): 2028–2036.
- 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

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

Van Den Bos, J., Rustagi, K., Gray, T., Halford, M., Ziemkiewicz, E., & Shreve, J. (2011). The \$17.1 billion problem: the annual cost of measurable medical errors. *Health Affairs*, 30(4): 596–603.

VI. Tables and figures

A. Tables

Statistic	2011	2010	2009	2008
Mean	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Standard deviation	0.01	0.01	0.02	0.02
Skewness	-0.13	-0.10	-0.09	0.03
Kurtosis	2.54	2.52	2.41	3.58

Table 1: Pridit summary statistics.

Notes: This table shows the summary statistics for the Pridit hospital scores from 2008-2011.

Measure type	Measure	Weighting	Rank
Demography	Emergency Department	0.24	33
Demography	Acute Care	0.43	25
Demography	Veterans Affairs	0.05	58
Demography	Non For Profit	0.16	37
Demography	For Profit	-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 Accreditation	0.28	31
Demography	DNV Accreditation	0.02	60

Measure type	Measure	Weighting	Rank
Demography	Number of Beds	0.72	12
Process: HA	Patients Given Aspirin at Arrival	-0.07	52
Process: HA	Patients Given Aspirin at Discharge	-0.05	56
Process: HA	Patients Given ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction	-0.09	51
Process: HA	Patients Given Smoking Cessation Advice/Counseling	0.01	69
Process: HA	Patients Given Beta Blocker at Discharge	-0.05	55
Process: HA	Patients Given Fibrinolytic Medication Within 30 Minutes Of Arrival	0.01	65
Process: HA	Patients Given PCI Within 90 Minutes Of Arrival	-0.02	61
Process: HF	Patients Given Discharge Instructions	0.11	43
Process: HF	Patients Given An Evaluation of Left Ventricular Systolic Function	0.27	32
Process: HF	Patients Given ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction	0.01	64
Process: HF	Patients Given Smoking Cessation Advice/Counseling	0.09	49
Process: PN	Patients Assessed and Given Pneumococcal Vaccination	0.10	46
Process: PN	Patients Whose Initial Emergency Room Blood Culture Was Performed Prior to the Administration of the First Hospital Dose of Antibiotics	0.00	70
Process: PN	Patients Given Smoking Cessation Advice/Counseling	0.13	38
Process: PN	Patients Given Initial Antibiotic(s) within 6 Hours After Arrival	-0.12	41
Process: PN	Patients Given the Most Appropriate Initial Antibiotic(s)	0.09	48
Process: PN	Pneumonia Patients Assessed and Given Influenza Vaccination	-0.01	68
Process: SCIP	Percent of surgery patients who were taking heart drugs called beta-blockers before coming to the hospital, who were kept on the beta-blockers during the	0.01	67

Measure type	Measure	Weighting	Rank
	period just before and after their surgery		
Process: SCIP	Surgery Patients Who Received Preventative Antibiotic(s) One Hour Before Incision	0.04	59
Process: SCIP	Percent of Surgery Patients who Received the Appropriate Preventative Antibiotic(s) for Their Surgery	-0.10	47
Process: SCIP	Surgery Patients Whose Preventative Antibiotic(s) are Stopped Within 24 hours After Surgery	-0.12	39
Process: SCIP	Cardiac Surgery Patients With Controlled 6 A.M. Postoperative Blood Glucose	-0.01	66
Process: SCIP	Surgery Patients with Appropriate Hair Removal	0.00	71
Process: SCIP	Urinary Catheter Removed on Postoperative Day 1 or Postoperative Day 2 with Day of Surgery being Day Zero	-0.12	40
Process: SCIP	Surgery Patients Whose Doctors Ordered Treatments to Prevent Blood Clots (Venous Thromboembolism) For Certain Types of Surgeries	0.02	63
Process: SCIP	Surgery Patients Who Received Treatment To Prevent Blood Clots Within 24 Hours Before or After Selected Surgeries to Prevent Blood Clots	-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

Measure type	Measure	Weighting	Rank
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
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 medications	-0.77	5
Satisfaction	Staff usually explained medications	0.36	30
Satisfaction	Staff gave recovery information	-0.41	26
Satisfaction	Hospital Rated 7-8 overall	0.62	20
Satisfaction	Hospital Rated 9-10 overall	-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 2: Pridit ranked variables (Lieberthal & Comer, 2013)

Notes: List of all hospital variables used for the Pridit analyses. Highlighted are the top ten variables impacting hospital Pridit scores.

Abbreviations- HA: heart attack, HF: heart failure, PN: pneumonia, SCIP: surgical care improvement, JC: The Joint Commission, ACGME: Accreditation Council for Graduate Medical Education, COTH: Council of Teaching Hospitals, DNV: DNV Healthcare Inc.

	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 3: Correlations of 2008-2010 Pridit scores and 2011 outcomes measures (Lieberthal & Comer, 2013)

Notes: Shows correlations of past hospital Pridit scores to the 2011 outcomes measures. This table demonstrates the use the past Pridit scores to predict future performance on outcomes measures.

B. Figures

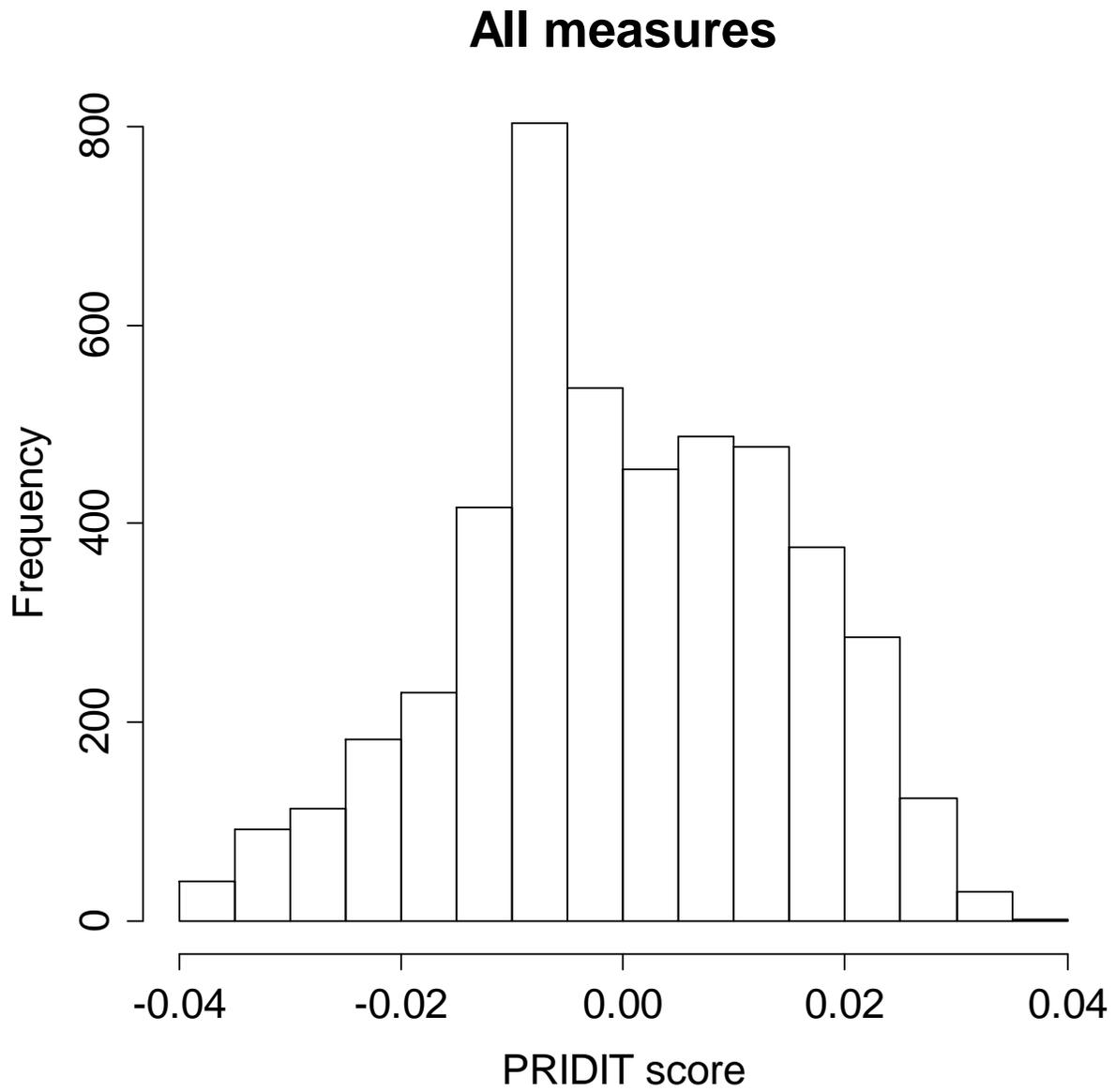


Figure 1: 2011 Pridit score distribution with all variables (Lieberthal & Comer, 2013)

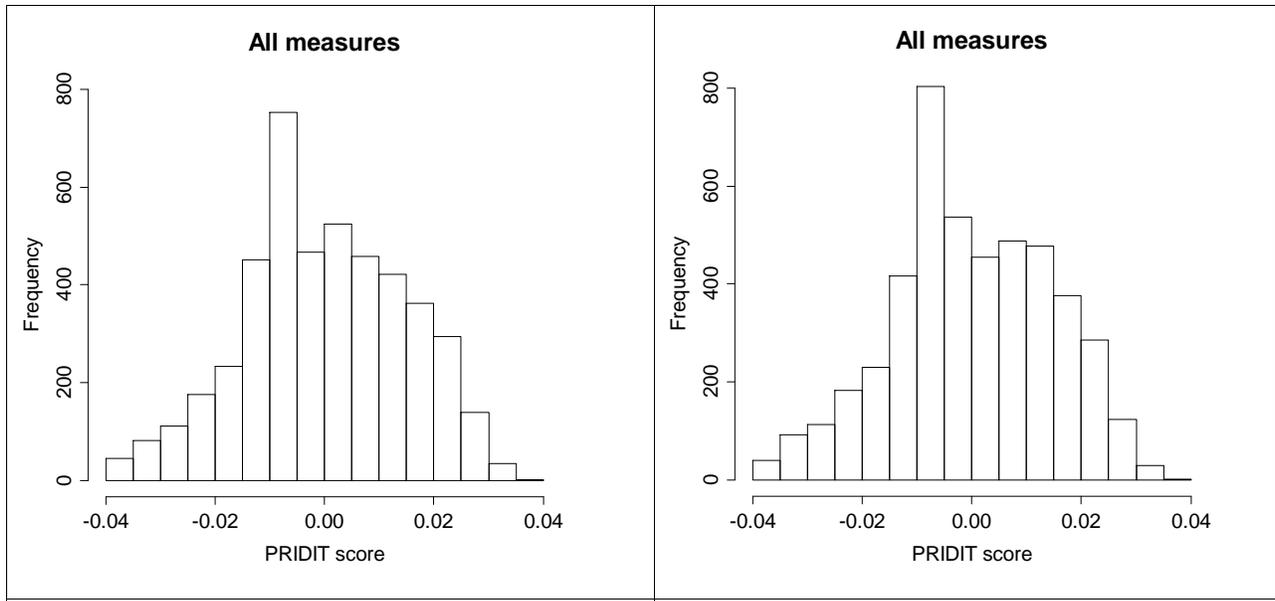


Figure 2: 2010 Pridit scores with all variables (Lieberthal & Comer, 2013)

Figure 3: 2011 Pridit scores with all variables (Lieberthal & Comer, 2013)

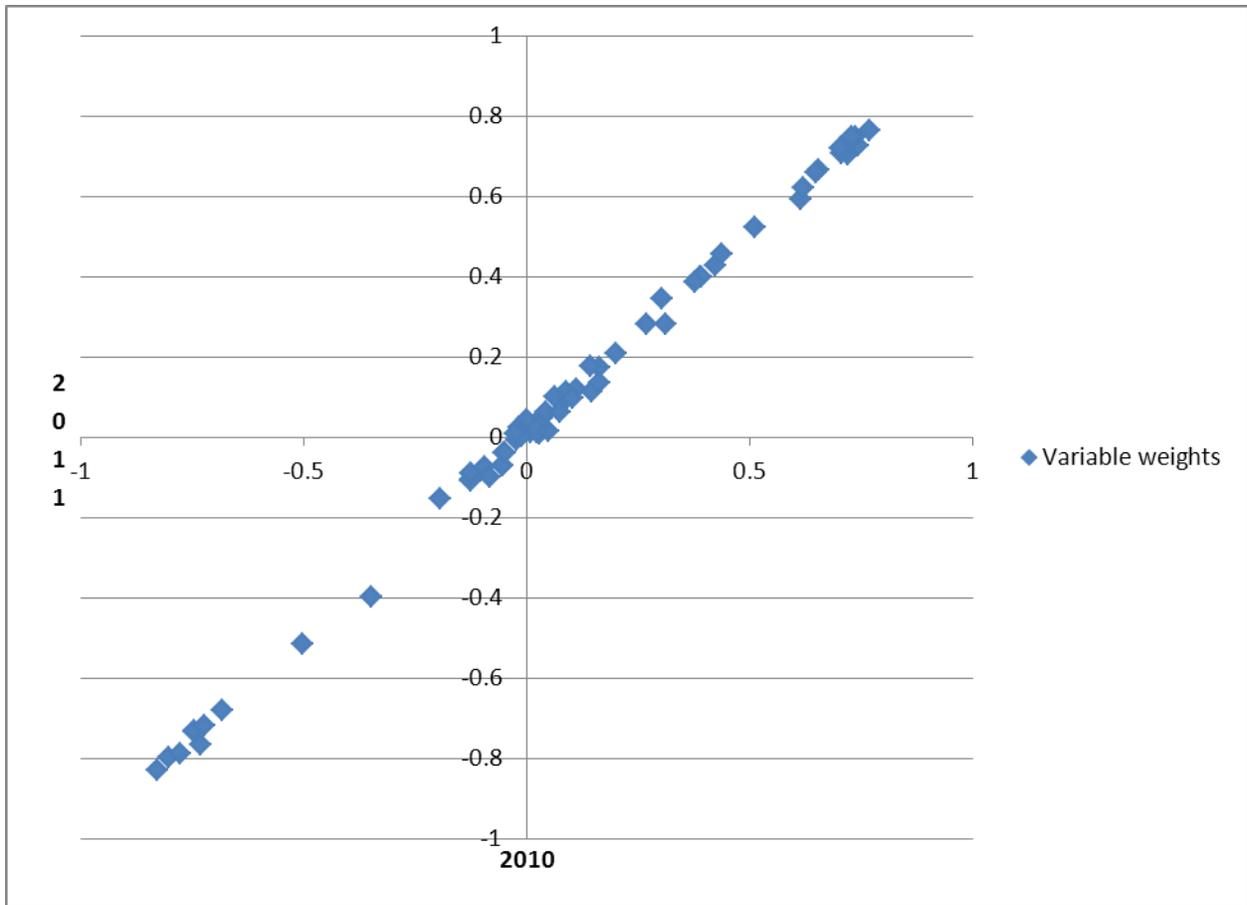


Figure 4: 2010-2011 variable weight comparison (Lieberthal & Comer, 2013)

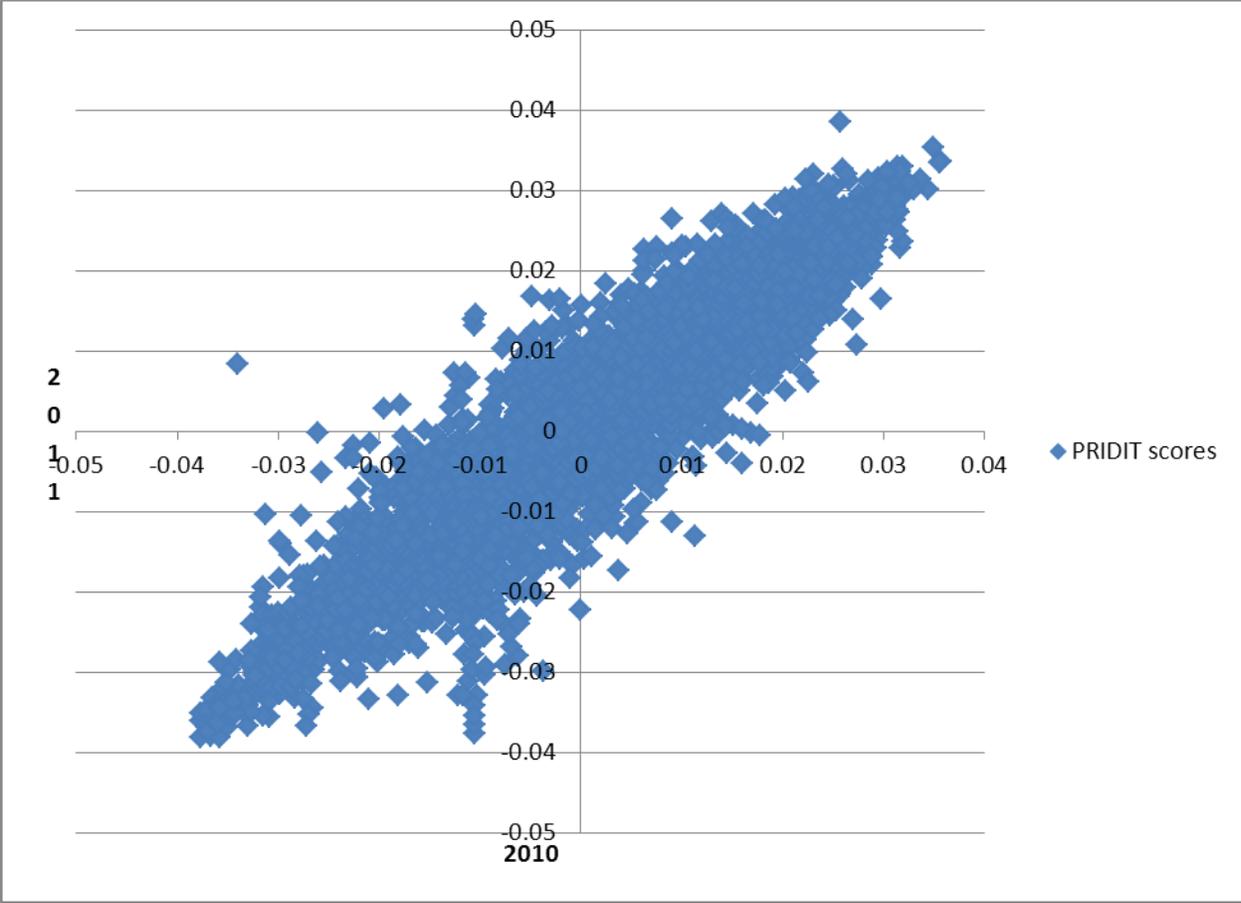


Figure 5: Correlation of 2010 and 2011 hospital scores (Lieberthal & Comer, 2013)