

INSIGHTS INSIGHTS INSIGHTS INSIGHTS INSIGHTS

Using Publicly Reported Data to Choose a Hospital:
Do Different Websites and Different Measures Give Consistent Answers?

Author: Steve Hines, PhD

December 17, 2003

INSIGHTS INSIGHTS INSIGHTS INSIGHTS INSIGHTS

I N S I G H T S



Abstract

Government and private websites are providing information to consumers about hospital characteristics and their performance on process measures and care outcomes such as risk adjusted readmission and mortality rates. The usefulness of all of this information to consumers depends on the extent to which different measures are consistent in identifying higher or lower performing hospitals. We evaluated the consistency of information about heart failure care in Maryland hospitals. Measures of utilization, processes, and care outcomes are available on websites created by Health Grades, CMS, and the Maryland Health Care Commission. Measures of volume were fairly consistent across the websites and the four heart failure process measures also were reasonably well aligned. However, readmission and mortality rates were negatively correlated with each other and essentially unrelated to the measures of processes and volume. These results suggest that consumers would have difficulty identifying good or bad hospitals for heart failure care based on current information on the web. They also point to the systematic need for further study of how publicly reported hospital data may impact consumer decision making, quality improvement activity, and processes used to collect and report performance information.

Introduction

Consumers are increasingly turning to the Internet to obtain information about medical conditions and their health care treatment. A variety of groups have responded to the demand for such information by constructing websites that provide information about hospital care for a variety of conditions. With only small difficulty, consumers can now access hospital mortality data on the Health Grades website¹, data on hospitals' care process measure performance at professional's site created by the Centers for Medicare and Medicaid Services (CMS)², and data on process measures and other outcomes on the websites of state health agencies such as the Maryland Health Care Commission (MHCC)³. These and other sites include utilization data that allows consumers to learn hospital size and the number of persons treated for different types of medical conditions.

A basic premise of each of these websites is that information on the site will allow consumers to make more informed choices about where to obtain care. More informed choices are viewed as beneficial by advocates of patient-centered medicine and by policy makers, who believe that consumers' ability to discriminate between higher and lower quality hospitals will ultimately force hospitals to improve the care they provide.

While hospital reporting websites reflect this shared premise, the types of information reported on these sites reveal differences in the assumptions they make about how consumers should determine which hospitals are better. Health Grades emphasizes mortality as a quality measure, while CMS's hospital reporting website stresses the importance of process measure performance, and MHCC's site includes information on process measures, readmission rates, and volume.

Experts would certainly disagree over the relative importance of these factors and details relating to whether measures were defined, measured, and risk adjusted appropriately. However, they would agree that a range of factors should be taken into account when determining which hospitals are better than others for care of a given condition. A consumer who reviewed the information on these websites would probably conclude that a better hospital for care of a condition is one that has higher volume, performs better on clinically-defensible process measures, and has better risk adjusted mortality and readmission rates.

Our Key Question

To date, no one has examined quality information provided across multiple websites to determine how consistent it is. This paper provides an initial answer to a very important question: Do the combined measures available across multiple selected websites provide consumers with consistent performance information? If some hospitals perform very well or poorly across different measures reported on different websites, then the value of these websites for facilitating informed consumer choice would be clear. On the other hand, if different measures and websites produce completely unrelated answers about which hospitals are "better," then proponents of hospital public reporting will need to carefully evaluate how their collective efforts are facilitating informed choice.

¹ <http://www.healthgrades.com/>

² <http://www.medicare.gov/Hospital/Home.asp?version=default&browser=IE%7C6%7CWin2000&language=English&defaultstatus=1&pagelist=Home>

³ <http://hospitalguide.mhcc.state.md.us>

Thus, the key question this paper addresses deals with the “alignment” of different quality measures rather than whether the measures are simply different on one website than another. If multiple measures and websites classify the same hospitals as “better” or “worse,” then consumers will be able to form stronger inferences than they could from any one measure or website.

Our Approach

For this paper, we further defined our key question by asking how consistent are utilization, process, and outcome measures for the treatment of heart failure within the state of Maryland. We chose Maryland because it is one of three states involved in the CMS Hospital Public Reporting project. Additionally, the MHCC contracted with Delmarva to produce and maintain a hospital reporting website that includes detailed information about Maryland hospital characteristics and process and outcome measures. Heart failure was selected as the clinical condition because all hospitals in Maryland are required to provide their results for the CMS/JCAHO heart failure process measures to MHCC and because Health Grades reports heart failure mortality rates for all Maryland hospitals. We also restricted our answer by limiting our data sources to three websites. These sources, as well as the measures from each that we included in our analyses, are summarized in Table 1.

Table 1. Measures and Data Sources for Analyses

<u>Measure Type and Name</u>	<u>Specific Measure</u>	<u>Data Source</u>
Utilization Measures:		
MHCC Volume	Total HF volume (CY 2002)	MHCC
HG Volume	Medicare HF volume (CY 2000-2002)	Health Grades
CMS Volume	Mean Denominator of 4 HF measures (10/02-3/03)	CMS/MHCC
Bed Size	Number of medical/surgical beds (2003)	MHCC
Process Measures (10/02-3/03):		
HF 1	Discharge Instructions	CMS/MHCC
HF 2	LVEF Assessment	CMS/MHCC
HF 3	ACE Inhibitor for low LVEF	CMS/MHCC
HF 4	Smoking Cessation Advice	CMS/MHCC
Outcome Measures:		
HG Mortality	Inpatient HF mortality (CY 2000-2002)	Health Grades
MHCC Readmit	14 day readmit rate for HF (CY 2002)	MHCC

Volume measures covered different time spans and somewhat different operational definitions of heart failure. Although the denominators for the heart failure process measures are not currently publicly reported, we included them in our analyses because they are available and may be publicly reported in the future. Because MHCC requires the reporting of the same process measures developed by CMS, we used data from the last quarter of 2002 and the first quarter of 2003 available on MHCC’s site rather than rates posted on the CMS website, which are from the second half of 2002. Health Grades provides actual and predicted mortality rates. We used this information to derive the risk adjusted mortality rates for our analyses. MHCC reports risk adjusted 14-day readmission rates, which we used in our analyses. Additional details regarding these measures or the analyses we performed are available on request from the author.

While Maryland has 47 acute care hospitals, we limited our analyses to those that had enough volume so that they could have a denominator of at least 20 for each of the ten measures summarized above. Once this set of 32 hospitals was identified, we performed the analyses summarized in Table 2.

Table 2. Selected Analyses

Purpose	Analysis	Variables
Assess the internal consistency of a combined volume measure	Cronbach's alpha	Four Volume measures
Assess the internal consistency of a combined process measure	Cronbach's alpha	Four Process measures
Examine the relationship between the two outcome measures	Pearson's correlation	Mortality and Readmission
Examine relations between Utilization, Process and Outcomes	Pearson's correlation	Combined volume and process measures and two outcome measures
	Regression analyses	Regressed four volume and four process measures on each outcome
Explore hospitals' consistency across utilization, process, and outcome measures	Grouped hospitals in 3rds; Classified hospitals based on groupings	Combined volume and process measures and two outcome measures

Our Findings

Consistency of volume measures: Table 3 reports the relationships between the four volume measures. Of the four measures, only the number of medical/surgical beds had weak associations with the other variables; the other three volume measures were strongly associated.

Table 3. Relations between Volume Measures

Pearson Correlation Coefficients and Probability levels, N = 32				
	HG Volume	Bed Size	MHCC Volume	CMS Volume
HG Volume	1.00	0.17 0.35	0.86 <.01	0.73 <.01
Bed Size	0.17 0.35	1.00	0.31 0.09	0.49 <.01
MHCC Volume	0.86 <.01	0.31 0.09	1.00	0.89 <.01
CMS Volume	0.73 <.01	0.49 <.01	0.89 <.01	1.00

The standardized Cronbach's alpha for the four-item volume measure was .84. The relatively strong associations between the three heart failure volume measures suggest that consumers can draw consistent conclusions across different websites regarding heart failure volume. Volume data from MHCC is from 2002, while Health Grades's volume data includes a three-year period ending in 2002. The CMS process measure volume data is from the last quarter of 2002 and the first quarter of 2003. This consistency of volume trends is clearly a benefit to consumers, who will always need to make inferences about what the hospital will be like in the future, based on data that is from the past.

For additional analyses using a combined volume measure, we used a measure derived from the average of the standardized scores for the four volume measures in Table 3. Hospitals were classified into three volume groups, with eleven in the higher volume group, ten in the lower volume group, and eleven in the average volume category.

Consistency of process measures: Table 4 reports the relationships between the four heart failure process measures. While not all the relationships are strong, all correlations between the measures are positive. The relations between HF 1 and HF 4 and HF 2 and HF 3 are .67 and .57, respectively. When the four process measures are combined, the resultant measure has a standardized Cronbach's alpha of .69.

Table 4. Relations between Heart Failure Process Measures

Pearson Correlation Coefficients and Probability levels, N = 32				
	HF 1	HF 2	HF 3	HF 4
HF 1	1.00	0.28 0.12	0.18 0.32	0.67 <.01
HF 2	0.28 0.12	1.00	0.57 <.01	0.22 0.23
HF 3	0.18 0.32	0.57 <.01	1.00	0.23 0.20
HF 4	0.67 <.01	0.22 0.23	0.23 0.20	1.00

These results suggest that a consumer could sensibly combine the four heart failure process measures into a single measure on which hospitals' performances could be compared. We created such a measure by averaging the standardized scores for each hospital on each of the four heart failure process measures. Hospitals were classified into three volume groups, with eleven in the higher and average process score groups, and ten in the low process score category.

Consistency of outcome measures: We correlated hospitals' risk adjusted inpatient mortality rate derived from the Health Grades website with their risk adjusted 14-day heart failure readmission rates reported by MHCC. This correlation was -.35, $p < .05$. While only moderate in size, the

direction of this relationship suggests that higher mortality rates may be associated with lower readmission rates. This negative relationship might be explained by hospitals' care for the sickest of their heart failure patients. Hospitals whose care restored function enough so that the patient could be discharged would probably have a higher readmission rate than hospitals that were not able to prevent the patient's death. Because the relationship between readmissions and mortality is complex, it is clear that they should not be combined into a single outcome measure. For our remaining analyses, we treated readmission and mortality rates as independent outcomes.

Consistency across volume, process and outcome measures: Table 5 shows the correlations between the combined volume and process measures and the two heart failure outcome measures. While it is true that the relatively small number of hospitals makes it harder to establish statistically significant relationships, only two of the relationships shown below are significant. Volume is totally unrelated to the heart failure process measure. Because hospitals of any size can have effective processes in place for heart failure care, this lack of relationship is unsurprising. Higher volume hospitals might be expected to have better mortality rates and lower readmission rates. Our results show a moderate negative relationship of $-.38$ between volume and readmission rates ($p < .04$). However, there was no relationship between the volume and mortality measures. Each heart failure process measure is clinically justified based on its association with better outcomes. While this would suggest that process rates might be negatively correlated with the mortality and readmission measures, we found nonsignificant positive correlations.

Table 5. Correlations between Volume, Process and Outcome Measures

Pearson Correlation Coefficients and Probability levels, N = 32				
	Combined Volume	Combined Process	MHCC Readmit	HG Mortality
Combined Volume	1.00	0.06 0.73	-0.38 0.04	0.10 0.58
Combined Process	0.06 0.73	1.00	0.08 0.65	0.13 0.47
MHCC Readmit	-0.38 0.04	0.08 0.65	1.00	-0.35 0.05
HG Mortality	0.10 0.58	0.13 0.47	-0.35 0.05	1.00

We performed two additional regression analyses to determine whether the four standardized volume and four standardized process measures could collectively predict readmission or mortality measure scores. These measures were not significant predictors of readmission scores ($p < .45$). And while they did predict mortality scores ($p < .04$, adjusted R-squared=.29) the nature of the relationship is complex. While some of the volume measures were positive predictors, others were

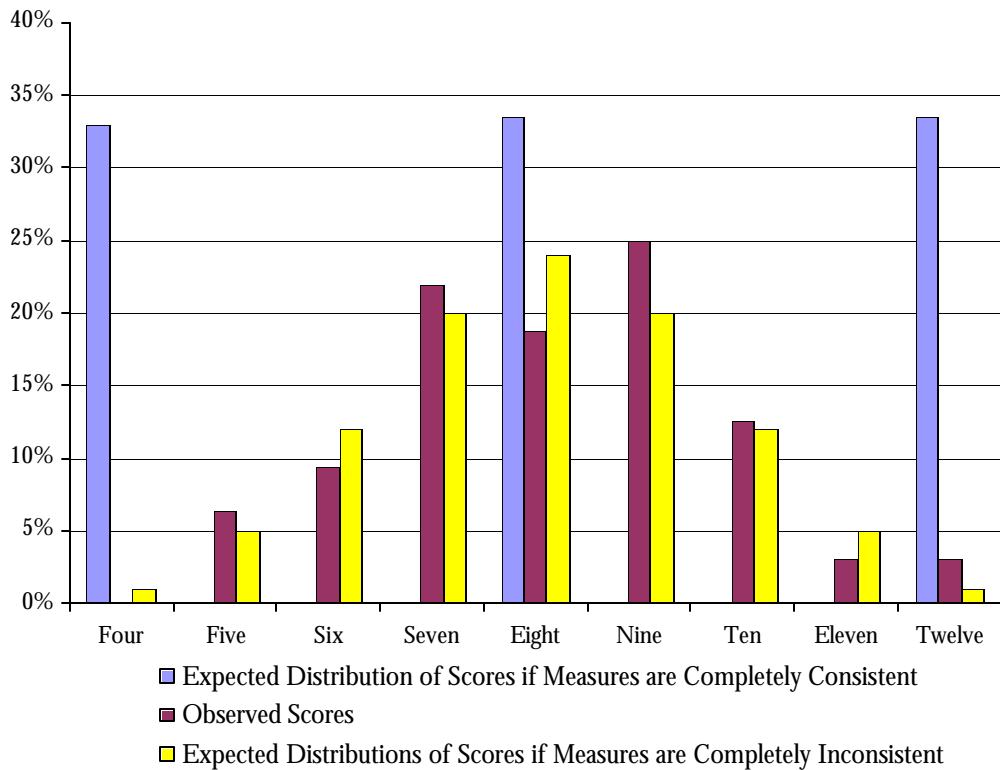
negative. The same blend of positive and negative predictors was observed for the four process measures. Thus, while it appears that more sophisticated analyses can establish relations between volume and process measures and publicly reported outcomes, these relations would be too complex to be of great benefit to a consumer attempting to select a hospital for heart failure treatment.

Consistency across grouped variables: Our final analysis grouped the hospitals into below average, average, and above average groups based on the combined measures of volume and process, and based on their mortality and readmission rates. The groups were formed by ranking hospitals from best to worst on each measure, and then placing the top eleven hospitals into the above average group, the middle eleven hospitals into the average group, and the bottom ten hospitals into the below average group. Because many hospital public reporting sites assign hospitals to three groups (although not necessarily into equal thirds), it is important to determine whether a consumer who used them could identify a hospital that scored consistently well or poorly across all four of these measures. To examine this possibility, we assigned a value of one if the hospital was below average on the measure, a value of two if it was average, and a value of three if it was above average. We then summed these four scores together to create a measure that could range from four (below average on everything) to twelve (above average on everything).

If the four measures were perfectly consistent indicators of quality, then there would be three groups of hospitals: the below average group (those with scores of four), the average group (with scores of eight), and the above average group (with scores of twelve). On the other hand, if the four measures were completely unrelated, then scores would have a bell shape, with very few extremely low or high scores and most scores between seven and nine.

Table 6 shows how the scores we observed compared to these two possible distributions. The observed distribution was significantly different from what would be expected if these four measures were completely aligned with each other ($p < .01$). But the distribution of scores was not significantly different from what would be expected if the four measures were completely unrelated to each other. Thus, although combining volume, process, and outcome measures will almost certainly result in several hospitals with very high and very low scores, this does not mean that the measures are related to each other. Instead, our analysis shows that consumers looking for the best and worst hospitals across these measures would draw conclusions more likely to be explained by chance than by any meaningful differences in the quality of heart failure care. It appears that the volume, process, and outcome measures we examined are not sufficiently aligned to provide direction to a consumer about where they should seek or avoid seeking care.

Table 6. Observed Grouped Scores Compared to Expected Scores if the Measures are Completely Related and Completely Unrelated



Conclusions

This paper provides a preliminary answer to what we believe will be an increasingly vital question for those involved in the public reporting of health care performance data. If multiple websites offer a variety of measures designed to help consumers select hospitals and other health care providers, we need to know whether this information is consistent. While well-aligned information would strengthen consumer confidence in using public reporting websites to make health care decisions, inconsistent information may actually call into question the value of all available data. Moreover, if hospitals and other care providers who monitor their performance find negligible relations between measures of processes and outcomes, they may question the value of efforts to improve process measure rates.

The analyses reported in this paper identify some areas where results are consistent, but reflect great inconsistency in other areas. Major conclusions include:

➤ **Utilization data is consistent across websites**

We found utilization data on hospital volume to be consistent across different populations (all payers versus Medicare versus persons included in the denominators for heart failure

process measure rates). This data also was consistent across multiple websites and across different time periods. Because utilization is more readily measured through administrative data this finding is unsurprising. However, it is still useful to know that reliance on a specific data source from a very recent time period may not be critical.

➤ **Process measures are consistent with each other**

We also found that specific heart failure process measure rates tend to be consistent with each other. This finding does not mean that hospitals cannot perform well on some indicators and poorly on others. But it does suggest that consumers may be able to form a reasonable global evaluation of the quality of important heart failure processes.

➤ **Utilization, process, and outcome measures are not consistent within or across sites**

When we compared process measures available on the CMS and MHCC websites with outcome measures reported by Health Grades and MHCC, we found negligible relationships. The failure to observe relations between processes and outcomes could easily be attributed to differences in time period, precise patient population, inappropriate measure definition or risk adjustment, and other equally plausible factors. And while sponsors of websites reporting hospital quality measures are likely to fault other websites for reporting inaccurate or invalid measures, such a response is unlikely to satisfy increasingly sophisticated health care consumers who recognize that no one measure or type of measure is adequate to reflect the overall quality of care a hospital provides.

Next Steps

We believe that additional study and monitoring of publicly reported hospital performance data is critical to improving the value of public reporting sites for consumers, facilitating quality improvement in hospitals, and optimizing the collection of data used in public reporting.

➤ **Implications for Hospital Public Reporting Website Development**

Additional study of publicly reported data is needed to continuously improve the value websites offered to consumers. Each public reporting website we examined offers value to consumers that goes beyond the data reported about each hospital. Understanding what information is most important to consumers for the selection of a hospital and for more informed discussions with health care providers is essential. Further study is also needed to develop more comprehensive measures of hospital quality that take into account a range of processes, outcomes, and other factors of concern to consumers. Without such study, consumers may either ignore available information because it is inconsistent, or they may rely on overly simplistic single measures, such as mortality, or overly simplistic aggregate measures such as those *US News and World Report* uses to rate hospitals.

➤ **Implications for Quality Improvement Activities**

A second reason for the importance of further study is its potential to impact quality improvement efforts in hospitals. Several studies have shown that hospitals monitor publicly reported data because they believe that consumers are monitoring it and they want to perform well. While the potential to use publicly reported data to motivate hospitals to commit to quality improvement projects is appealing, there are risks as well. If hospitals believe consumers are more interested in reported mortality rates than process rates, changing mortality rates may become a greater priority than process-linked improvement efforts. And if they believe that process measures are unrelated to outcomes, this belief may reduce resources they will commit to strengthening those processes. Alternatively, if further study can establish relations between process measures and specific outcomes, new opportunities may exist for improving care quality.

➤ **Implications for Data Collection Processes**

A final reason that further study is critical is because such study is needed to improve the efficiency of ongoing efforts to collect and publicly report hospital quality data. For example, a better understanding of how long it takes for a hospital's rates to change on particular quality indicators would allow for data-based decisions regarding how frequently data must be collected, what time periods rates should be reported for, and how often publicly reported data should be updated. Further study of how process variables relate to each other can also lead to statistically-based methods for monitoring and validating data at both the hospital and state levels. If moderate positive associations between process measures prove to be the norm, then hospitals or geographic areas that fail to reflect this pattern could receive further scrutiny.

While future research could go in many directions, important next steps need not be highly complex. Determining whether measures are stable across time is very important to consumers who need to know whether published rates are valid predictors of the care they might seek in a later period. Learning whether the patterns that we observed for Maryland heart failure care are consistent across different clinical areas and different geographic regions are equally important. Data to perform both these types of analyses is available now or will be available soon. Additional analyses of these types will make an important contribution to the challenging and multidimensional effort to improve the quality of care hospitals provide.

Acknowledgments:

Steve Hines (PhD, Purdue University) is a Senior Scientist at the Delmarva Foundation. Contributions to the development, analysis and writing of this paper from David Boan, Dianne Feeney, Matthew Fitzgerald, Frank Funderburk, Maulik Joshi, Mariana Leshner, and Kathryn Schulke are gratefully acknowledged. Additional information regarding this paper is available from the author at hinesst@dfmc.org. Information on other papers produced as part of the Delmarva Insight Series is available from Kathryn Schulke at kschulke@dfmc.org. Although the Delmarva Foundation has contracts with CMS and MHCC, no funds from these contracts were used to support the development of this paper, which represents the independent assessment of Delmarva.