Improving Medicare’s Hospital Compare Mortality Model

Jeffrey H. Silber, Ville A. Satopää, Nabanita Mukherjee, Veronika Rockova, Wei Wang, Alexander S. Hill, Orit Even-Shoshan, Paul R. Rosenbaum and Edward I. George

**Objective.** To improve the predictions provided by Medicare’s Hospital Compare (HC) to facilitate better informed decisions regarding hospital choice by the public.

**Data Sources/Setting.** Medicare claims on all patients admitted for Acute Myocardial Infarction between 2009 through 2011.

**Study Design.** Cohort analysis using a Bayesian approach, comparing the present assumptions of HC (using a constant mean and constant variance for all hospital random effects), versus an expanded model that allows for the inclusion of hospital characteristics to permit the data to determine whether they vary with attributes of hospitals, such as volume, capabilities, and staffing. Hospital predictions are then created using directly standardized estimates to facilitate comparisons between hospitals.

**Data Collection/Extraction Methods.** Medicare fee-for-service claims.

**Principal Findings.** Our model that included hospital characteristics produces very different predictions from the current HC model, with higher predicted mortality rates at hospitals with lower volume and worse characteristics. Using Chicago as an example, the expanded model would advise patients against seeking treatment at the smallest hospitals with worse technology and staffing.

**Conclusion.** To aid patients when selecting between hospitals, the Centers for Medicare and Medicaid Services (CMS) should improve the HC model by permitting its predictions to vary systematically with hospital attributes such as volume, capabilities, and staffing.

**Key Words.** Medicare quality of care, Bayesian statistics, hospital compare, acute myocardial infarction

There are many organizations currently creating public reports on hospital quality (ConsumerReports.org 2014; New York State Department of Health 2014; U.S. News & World Report 2014; Healthgrades.com 2015; Scientific Registry of Transplant Recipients), but a leader in this endeavor is Medicare’s Hospital Compare (HC; Krumholz et al. 2006a,b,c; U.S. Department
HC is important, not only because patients, caregivers and administrators use the website but also because the methodology chosen by HC is often adopted by other organizations. For example, the Society of Thoracic Surgeons and Consumer Reports use a model patterned after the HC mortality model (O’Brien et al. 2009; The Society of Thoracic Surgeons 2015), and Consumer Reports also uses results of the HC model in parts of their safety metric (Consumer Reports Health 2014). The problem with this “follow-the-leader” approach to public reporting in health care is that there is still controversy regarding such methods (Mukamel et al. 2010; Silber et al. 2010; Jones et al. 2014) because of Medicare’s decision not to utilize hospital characteristics when making its predictions. Despite attempts to help settle some issues surrounding the hierarchical random effects model utilized by HC, such as the 2011 report produced by the Committee of the Presidents of the Statistical Societies (Ash et al. 2012), the HC model continues to not include hospital characteristics (U.S. Department of Health & Human Services. Centers for Medicare & Medicaid Services. Centers for Medicare & Medicaid Services).

This report examines the above-mentioned controversy by reframing the HC model in a Bayesian perspective. The HC random effects model makes powerful assumptions about the nature of hospital outcomes—it assumes that outcomes vary among hospitals for two reasons: differing patient populations and a random hospital effect that is implicitly independent of attributes of hospitals such as volume (Cameron and Trivedi 2005; Normand and Shahian 2007). We offer an alternative model that tests whether or not hospital...
quality varies with such things as volume and staffing. We will (1) briefly describe a Bayesian model that mimics the HC model; (2) describe an expanded Bayesian model that includes hospital characteristics thereby allowing the data to determine if hospital outcomes vary predictably with these hospital characteristics; (3) make an argument for using direct standardization rather than the indirect standardization method of reporting utilized by HC to compare these hospitals; and (4) compare hospitals in Chicago that saw Medicare patients with acute myocardial infarction using two Bayesian models, HC’s model without hospital characteristics and our expanded model that includes hospital characteristics.

METHODS

Patient Population

We obtained the Medicare claims data on patients admitted with acute myocardial infarction (ICD9 codes 410, 4100, 41000, 41001, 4101, 41010, 41011, 4102, 41020, 41021, 4103, 41030, 41031, 4104, 41040, 41041, 4105, 41050, 41051, 4106, 41060, 41061, 4107, 41070, 41071, 4108, 41080, 41081, 4109, 41090, 41091) for patients 65 and older from the entire United States for the years 2009, 2010, and 2011. Databases included the inpatient claims (MEDPAR), outpatient claims, Physician/Part B claims, and Hospice claims. To construct the dataset, we followed closely the methods described by Krumholz and Normand (Krumholz et al. 2006b,c), and those described by HC (U.S. Department of Health & Human Services, Centers for Medicare & Medicaid Services). For each year 2009–2011, the denominator file was appended to that year’s MedPAR data by beneficiary ID, allowing date of death to be tracked inside or outside the hospital. Only admissions to acute care hospitals were used, as identified by the provider ID. Patients who were in an HMO within 180 days prior to, or 30 days following discharge were excluded. Patients under the age of 65 were also excluded. The index admissions were restricted to those with admission dates between July 1, 2009 and December 31, 2011, and only one randomly selected admission per patient was allowed. If the admission date fell between July 1, 2009 and June 30, 2011, the record was assigned to the development sample; if the admission date fell between July 1, 2011 and December 31, 2011, the record was assigned to the validation sample. There were 364,677 unique patients treated at 4,396 hospitals.
Model Construction

The original HC model was fit using a program \textit{SAS GLIMMIX} (SAS Institute 2008) that can only fit a limited set of models. In contrast, the Bayesian approach is more flexible, and one can fit a model with less attention to limitations imposed by the software (Zhao et al. 2006), while at the same time being able to take advantage of the many approaches to describing model results that are commonly used in the Bayesian framework (Gelman et al. 1997a). We will fit random effects models using a Bayesian framework similar to Austin, Naylor, and Tu (2001) and Normand and Shahian (2007). Other approaches and suggested approaches are in the literature (Normand, Glickman, and Gatzonis 1997; Ohlssen, Sharples, and Spiegelhalter 2007; Ash et al. 2012; Jones et al. 2014). We start by fitting the original HC model using Bayesian methods, and then expand the model to permit it to resemble unambiguous patterns in Medicare data.

To describe public reporting models, we use the following notation: for 30-day mortality of the \(i\)th patient in hospital \(h\), \(Y_{hi} = 1\) if dead or \(Y_{hi} = 0\) if alive; for attributes of patients upon, or prior to, admission such as age or diabetes, \(x_{hi}\); for attributes of hospital \(h\) such as volume, \(z_h\). Hospital \(h\) provides data on patients \(i = 1, 2, \ldots, n_h\) for \(h = 1, \ldots, H\). The HC model introduces a random hospital effect \(\alpha_h\) for each hospital \(h\), that effect \(\alpha_h\) being assumed to be sampled from a normal distribution with constant mean and constant variance independent of all attributes of hospitals, \(\alpha_h \overset{iid}{\sim} N(\mu, \sigma^2)\). The HC model is generally written as follows:

\[
\Pr(Y_{hi} = 1|\alpha_h, x_{hi}) = p_{hi} \text{ with conditional independence of the } Y_{hi}
\]

(1)

\[
\log\{p_{hi}/(1 - p_{hi})\} = \text{logit}(p_{hi}) = \alpha_h + \beta^T x_{hi}
\]

(2)

\[
\alpha_h \overset{iid}{\sim} N(\mu, \sigma^2)
\]

(3)

see Normand and Shahian (2007) (their equations (4) and (5) on p. 214). Normand and Shahian (2007, p. 214) then correctly observe: “An implicit assumption in the model defined by (their equations (4) and (5)) is that hospital mortality is independent of the number of patients treated at the hospital.” Let us make this implicit assumption explicit by writing the model as:

\[
\Pr(Y_{hi} = 1|\alpha_h, x_{hi}, z_h) = p_{hi}
\]

(1’)

with conditional independence of the \(Y_{hi}\) (1’)

\[
\log\{p_{hi}/(1 - p_{hi})\} = \text{logit}(p_{hi}) = \alpha_h + \beta^T x_{hi}
\]

(2’)

4 HSR: Health Services Research
Whether one writes (1)–(3) or (1′)–(3′), the same estimates and confidence intervals are produced. The distinction is that (1′)–(3′) states the implicit assumption explicitly, namely the hospital attributes \( z_h \) such as volume, are not related to \( x_h \). It is important to realize that the HC model is a model, and like any scientific model, it might be correct or incorrect, and the advice received from such a model might be useful or misleading.

Once we have a model for the patient mortality probabilities, \( p_{ih} \), we also have a model for the hospital mortality rates:

\[
p_h = \left( \sum_i p_{hi} \right) / n_h
\]

The HC model (1′)–(3′) says hospital mortality rates vary from one hospital to another because they see different patients, reflected in \( x_{hi} \), and because they have different quality parameters \( x_h \) that were drawn randomly from a single normal distribution.

Details of specification and fitting for the Bayesian models are provided in the Statistical Appendix, but briefly, we add relatively noninfluential neutral prior distributions for all the unknown parameters and then use Markov Chain Monte Carlo (MCMC) posterior simulation for calculations (for more details, see George et al. 2015).

A different “expanded” model permits hospital quality \( x_h \) to be related to hospital attributes \( z_{ih} \) such as volume, by allowing the expectation of \( x_h \) given \( z_h \) to be a function of \( z_{ih} \): \( E(x_h|z_h) = \mu(z_h) \):

\[
x_h|z_h \sim N\{\mu(z_h), \sigma^2\}
\]

It is important to realize that this expanded model permits hospital quality to change with hospital volume (or other hospital attributes). This expanded model permits \( E(x_h|z_h) = \mu(z_h) \) to vary with \( z_{ih} \) but it does not require it to do so.

The hospital characteristics we consider are as follows: hospital volume, an indicator for whether a hospital performs percutaneous transluminal cardiac angiography, stenting, or coronary artery bypass surgery, which we will call percutaneous coronary intervention or “PCI” for short; the hospitals resident-to-bed ratio (RB ratio); and finally the hospitals nurse-to-bed ratio (nurse-to-bed [NTB] ratio). Hospital volume was defined as the total volume of admission for the study conditions noted above over the 3 years of the CMS dataset, but unlike in the fitting of our models, we allowed patients to be
counted multiple times if they had multiple admissions. Each hospital’s resident-to-bed ratio and NTB ratio and number of beds were determined using the Medicare Provider of Service file. NTB ratio was defined by dividing the number of full-time-equivalent registered nurses and licensed practical nurses by the number of total beds. Likewise, resident-to-bed ratio was defined by dividing the number of residents by the number of total beds.

To us, it is an empirical matter whether hospital quality $z_h$ is unrelated to hospital attributes $z_h$, as in (3’), or whether quality is related to attributes, as in (3’’) with $E(x_h|z_h) = \mu(z_h)$. Normand, Glickman, and Gatsonis (1997, p. 812) agree: “consideration of provider characteristics as possible covariates is dictated by the need to explain as large a fraction as possible of the variability in the observed data. Simple exchangeability across all providers may not be a defensible assumption for many datasets.” Consistent with this view, Cameron and Trivedi (2005, p. 701) say that random effects models that force $E(x_h|z_h)$ to be constant and not varying with $z_h$, as in (3’), are often not appropriate for microeconomic analysis. Our model permits the conditional expectation $\mu(z_h)$ of $x_h$ to vary with volume in accordance with a spline function (see George et al. 2015). The other hospital characteristics $z_h$ enter $E(x_h|z_h) = \mu(z_h)$ linearly; only the volume enters as a spline.

The difference in assumptions between (3) and (3’’) can have a great effect on the estimates of hospital quality, $x_h$. This has been illustrated in the context of predicting mortality in AMI patients (Silber et al. 2010).

Methods to Display Model Predictions

Combing the data with Bayes specifications of each of the models produces posterior (post-data) probability distributions of mortality rates for every hospital. In this report we will display the results of the models in two ways. First, we will provide posterior probability density plots for mortality rates at two Chicago hospitals, one small and one large. Next, for five hospitals in Chicago (and in the Electronic Appendix, all hospitals in Chicago), we provide the posterior probability that a specified hospital $h$ has a higher mortality component $x_h$ than another specified hospital (suggesting which hospital in the pair should be avoided).

For each model, these results were obtained from an MCMC-generated sample of 10,000 random draws from the posterior distribution of the model parameters $(x_h, \beta)$. All the Bayesian models were estimated with Gibbs sampling (Geman and Geman 1984; Gelman et al. 1997c). These samplers have been written in the C++ programming language with all necessary operators
deriving from either the Standard Library (ISO 2013) or the linear algebra library called Armadillo (Sanderson 2010).

One other aspect of this report concerns the comparison of hospitals using direct standardization by creating a counterfactual table of predictions, one row for every patient in the population, one column for every hospital, the table entry being the probability that the patient in this row would die if treated at the hospital in this column. The hypothetical here is that all patients in the dataset went to every hospital. In the column for hospital $h$, each entry is then the patient’s probability of dying using the $x_h$ of hospital $h$, but using their specific $x_{hi}$ characteristics. The average mortality rate for this column, denoted $P^D_{h}$, defines the directly standardized probability of mortality at hospital $h$. Each Bayes model provides a posterior distribution for every entry of this table and hence for every $P^D_{h}$.

Consider one of the 10,000 draws ($\alpha$ and $\beta$). The direct standardization approach computes:

$$P^D_{h} = (1/N) \sum_j \logit^{-1}(\alpha_h + \beta x_{ji})$$

where $P^D_{h}$ is approximately the directly standardized probability of mortality, $N$ is the total number of all patients across all hospitals ($N = \sum h n_h$), and the $\alpha$’s and $\beta$’s are parameters in a logit function that are fixed over all patients. This is repeated 10,000 times to get 10,000 different draws of $P^D_{h}$. This forms our posterior sample of $P^D_{h}$.

In this report we will display directly standardized plots of hospital posterior probability density functions for 30-day mortality under the HC model (without hospital characteristics) and then our expanded model.

**RESULTS**

We first examine predictions from the Bayesian HC model that was constructed to resemble the HC model developed using GLIMMIX, the methodology of HC. To provide an idea of how similar the predictions are in both models, we compare the predicted probabilities of 30-day mortality for each hospital using the HC Bayesian model we derived, with those from the GLIMMIX fit of the HC model on the same data from 4,396 hospitals. The correlation of the hospital mortality estimates was 0.997, the median absolute error was only 0.11 percent, and a maximum error was only 0.79 percent. As
the Bayesian HC model so closely mimics the HC GLIMMIX model, going forward, we will only discuss the Bayesian models.

In Table 1, we describe five hospitals in Chicago with respect to the variables utilized in the expanded models that include hospital characteristics to be compared to HC’s model that does not include hospital characteristics. Each hospital is described by its volume (and we will identify and refer to each hospital by its volume), as well as its ability to perform cardiac procedures, as defined by our PCI variable, which reflects technologies useful for patients presenting with acute myocardial infarction, as well as the hospital’s NTB ratio and resident-to-bed ratio. For purposes of clarity, we describe only five hospitals, but in the online appendix we present results for all 27 hospitals in Chicago.

We next examine the relationship between the Bayesian HC model and the expanded model including hospital characteristics, using direct standardization. A hospital with a small size \((n = 17)\) is illustrated in Figure 1A, which examines the difference between the two models using direct standardization and suggests that the model including hospital characteristics predicts far higher mortality than the model that does not include such characteristics. We display a larger hospital \((N = 448)\) in Figure 1B. Here we see that there is almost complete overlap in posterior probability density functions, suggesting less difference in predictions in the large hospital, in part because there is less shrinkage for both models (due to the large size).

The comparison of five Chicago hospitals using direct standardization \((P_h^{DS})\) is shown in Table 2 that provides the predicted mortality rates using direct standardization. Figure 2 displays these five hospital predicted mortality rates for both the HC model in Figure 2A and the expanded model that includes hospital characteristics in Figure 2B. As can be seen, using the model without hospital characteristics, we observe one very good hospital (a large hospital of volume 448), with predicted mortality rates better

Table 1: A Description of Five Hospitals in Chicago That Treat Medicare Patients with Acute Myocardial Infarction

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Volume</th>
<th>Performs PCI</th>
<th>Nurse-to-Bed Ratio</th>
<th>Resident-to-Bed Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>17</td>
<td>0</td>
<td>2.680</td>
<td>0.000</td>
</tr>
<tr>
<td>D</td>
<td>26</td>
<td>0</td>
<td>0.665</td>
<td>0.017</td>
</tr>
<tr>
<td>F</td>
<td>44</td>
<td>0</td>
<td>2.410</td>
<td>0.040</td>
</tr>
<tr>
<td>U</td>
<td>313</td>
<td>1</td>
<td>1.915</td>
<td>0.648</td>
</tr>
<tr>
<td>Y</td>
<td>448</td>
<td>1</td>
<td>2.498</td>
<td>0.569</td>
</tr>
</tbody>
</table>
(lower) than the four other hospitals, all these with mortality rates around 16 percent (slightly worse than the national average of 14.9 percent). However, using the expanded model that includes hospital characteristics, we see a different result. The moderately large hospital of volume 313 is now predicted to perform better than the three smallest hospitals, and the three smallest hospitals all have shifted to considerably higher predicted mortality rates.

To further examine how different the models may be in comparing these five hospitals, we report in Table 3 all pairwise comparisons of predicted hospital comparisons using a directly standardized approach for both the HC and the expanded model including hospital characteristics (E). As can be seen, small hospitals in the HC model all have probabilities of higher mortality compared to other hospitals in Chicago, of around 0.5. In other words, for small hospitals using the HC model, we generally can consider these across hospital differences a toss-up with respect to mortality. The story is very
Table 2: Posterior Probability Distributions of Five Hospitals in Chicago under Two Models, the Hospital Compare Model (without Hospital Characteristics) and the Expanded Model (Including Hospital Characteristics of Volume, Nurse-to-Bed Ratio, Resident-to-Bed Ratio, and the Ability to Perform Cardiac Catheterization Procedures)

<table>
<thead>
<tr>
<th>Hospital Volume</th>
<th>Expanded Model (with hospital characteristics)</th>
<th>Hospital Compare Model (without hospital characteristics)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Credible Interval*</td>
<td>Credible Interval</td>
</tr>
<tr>
<td></td>
<td>2.5th% 97.5th% Mean</td>
<td>Mean 2.5th% 97.5%</td>
</tr>
<tr>
<td>17</td>
<td>14.5 24.7 19.3</td>
<td>15.0 10.6 20.2</td>
</tr>
<tr>
<td>26</td>
<td>15.0 25.0 19.7</td>
<td>15.8 11.4 21.0</td>
</tr>
<tr>
<td>44</td>
<td>14.2 23.9 18.7</td>
<td>16.7 12.0 22.0</td>
</tr>
<tr>
<td>313</td>
<td>11.9 19.5 15.4</td>
<td>16.3 12.4 20.8</td>
</tr>
<tr>
<td>448</td>
<td>8.6 14.0 11.1</td>
<td>11.1 8.4 14.1</td>
</tr>
</tbody>
</table>

Notes. In the expanded model, the credible interval for Hospital 448 did not overlap with Hospitals 17, 26, and 44. For the Hospital Compare model, all hospitals had overlapping credible intervals.

*Unlike the more familiar 95% confidence interval, the 95% credible interval for the mortality rate is an interval that contains the true rate with 95% posterior probability.

different when we observe the model that includes hospital characteristics. We now see that the small hospitals generally are almost always worse than the larger hospitals.

DISCUSSION

Medicare’s HC website utilizes indirect standardization for public reporting by examining the predicted rate of mortality over the expected rate of mortality at each hospital, where “predicted” refers to a prediction made by the HC random effects model (a model that does not include hospital characteristics) and “expected” is derived from a model that utilizes only patient characteristics in the model. This “P/E” framework replaced the standard observed over expected or “O/E” approach utilized by most report cards prior to HC’s introduction (Silber et al. 2010). Replacing the actual observed outcome with the predicted was motivated by the desire to stabilize the observed rates, especially when hospitals are small, and observed rates are unstable. In estimating each hospital’s predicted mortality rate, the HC random effects model presently includes no hospital characteristics, instead using only patient
characteristics and a random effect for the hospital. The “E” for both the HC “P/E” and traditional indirect standardization approach using “O/E” utilizes only patient characteristics. In the expanded model we presented in this report, the “P” now includes both patient and hospital characteristics, allowing

Figure 2: Plots Comparing Five Chicago Hospitals under the Hospital Compare Model (A above) and under the Expanded Model That Includes Hospital Characteristics (B below) Using the Directly Standardized Approach

(A) Hospital Compare directly standardized mortality rates ($P_{DS}$)

(B) Expanded directly standardized mortality rates ($P_{DS}$)
for a better estimate of the predicted mortality rate for each hospital. The E of the expanded model’s “P/E” also only utilizes patient characteristics.

In this report, we recast Medicare’s HC model within a Bayesian framework. This allowed us to examine properties of the present HC model and compare it to a model that allows hospital characteristics to enter into the prediction. We find that by allowing a model to fit coefficients regarding hospital characteristics and mean mortality rates, instead of making assumptions that implicitly suggest that such hospital characteristics are uncorrelated with outcome, the expanded model suggests that hospital characteristics including volume and other variables such as NTB ratio and resident-to-bed ratio do matter. In the HC model, all hospitals are derived from a single overall distribution with one constant for the mean of all hospitals. In our model, we could have found that the HC assumption best fit the observed data, but we did not. Instead, our Bayesian expanded model approach that included hospital characteristics provided different results than the HC model. In our other work (George et al. 2015), we formally describe how much better various models that include hospital characteristics perform in terms of out-of-sample Bayes factors (Gelman et al. 1997b), and we begin to explore assumptions on the variance as well. An expanded model that includes hospital volume displayed an out-of-sample log Bayes factor of 31.93 compared to the baseline HC model, suggesting a vast improvement in model fit, and an even larger log Bayes Factor (the model we described in the present report) when also adding in the three hospital characteristics (log Bayes Factor = 34.58), again using the HC model as the base case.

Table 3: Pairwise Comparisons for Selected Hospitals in Chicago Using Direct Standardization

<table>
<thead>
<tr>
<th>HOSP N</th>
<th>17 HC</th>
<th>17 E</th>
<th>26 HC</th>
<th>26 E</th>
<th>44 HC</th>
<th>44 E</th>
<th>313 HC</th>
<th>313 E</th>
<th>448 HC</th>
<th>448 E</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>.40</td>
<td>.46</td>
<td>.31</td>
<td>.56</td>
<td>.34</td>
<td>.88</td>
<td>.92</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>.60</td>
<td>.55</td>
<td>.41</td>
<td>.61</td>
<td>.44</td>
<td>.91</td>
<td>.96</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>.69</td>
<td>.44</td>
<td>.59</td>
<td>.39</td>
<td>.55</td>
<td>.86</td>
<td>.98</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>313</td>
<td>.66</td>
<td>.12</td>
<td>.56</td>
<td>.09</td>
<td>.45</td>
<td>.14</td>
<td>.98</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>448</td>
<td>.08</td>
<td>.00</td>
<td>.04</td>
<td>.00</td>
<td>.02</td>
<td>.00</td>
<td>.02</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. The probability that the hospital defined by the row has a greater predicted mortality rate than the hospital defined by the column (see appendix for all pairwise results for Chicago hospitals). For each hospital defined in the column, we present the Hospital Compare (HC) Hospital estimate and then the expanded (E) estimate that includes hospital characteristics. Note that for smaller hospitals, the HC and E results often display quite different predictions, with the HC model frequently producing a “toss-up” value near 0.5, while the expanded model that includes hospital characteristics displays far more extreme probabilities, implying one hospital is very different from the other.
The purpose of this report was not to make a definitive list of hospital characteristics that should be utilized by HC. Instead, it was to definitively show that adding hospital characteristics into the HC model will produce a better model, providing the public with very different results. One may ask the question, when is using a random effects model appropriate for public reporting? Our perspective in this report, and in our earlier work (Silber et al. 2010) has been from that of the patient. The rationale for using the expanded model presented in this study is that patients will be provided with better predictions of mortality to make better choices for receiving care. In the expanded model, hospital characteristics that generally indicate better mortality (say PCI or increased volume) can be utilized to direct patients away from specific hospitals that do not perform PCI and have small volume. If patients instead utilized the HC model, which does not include hospital characteristics, they would not be directed away from these hospitals. While there may be some small hospitals with excellent outcomes despite not performing PCI, the vast majority of such hospitals perform worse than those larger hospitals that do perform PCI. On average, the public would be better directed to the superior hospitals and better able to avoid poor hospitals, when the HC model is expanded to include hospital characteristics. At the same time, for a hospital administrator working in a small hospital without PCI, improvements in outcomes may not be reflected in the expanded model’s predictions. We would suggest that neither the HC random effects model nor the expanded model may be appropriate to assess that small hospital administrator’s worth.

One other way to help the public compare hospitals using the HC website is to move away from the “P/E” approach and present the directly standardized posterior probability density of each hospital for any outcome of interest, and also to present, as we describe with 30-day mortality, the probability that one hospital is performing worse (or better) than another. Such computations can readily be performed in a Bayesian context, as we have described herein. Given that HC has decided to use a Random Effects model, we would recommend that Medicare move to recast their model development and presentation for public reporting in a Bayesian manner, as we have presented in this report. Such presentations can be as simple as generating tables similar to Tables 2 and 3, and generating graphs similar to Figure 2.

The most important finding of this report is that the predicted mortality rates were very different when using a model that does not include hospital characteristics versus a model, especially for small hospitals, that does include hospital characteristics. Not only was this very apparent when viewing Figure 2 but also when we examined credible intervals in Table 2 and
probabilities of worse outcomes in Table 3. If patients had to choose between two hospitals, and wished to avoid the hospital with the higher predicted mortality, it was apparent that the model with hospital characteristics gave very different advice than the HC model. Specifically, in the HC model, all hospital credible intervals overlapped at least one other hospital. This was not true when using the model with hospital characteristics. Such findings were consistent with previous work by Austin et al. (2004) examining acute myocardial infarction admissions from Ontario hospitals. They found that case volume and other hospital characteristics did influence the random effects model mortality rate predictions.

There is a common misconception surrounding the random effects model used by HC, resulting in the incorrect belief that all hospital characteristics are already included in the model through the hospital identifier and therefore negating the need for including hospital characteristics in the model. As the argument goes, as the random effects model includes a hospital identifier, the predictive model will account for all its characteristics just like the well-known fixed effects model will. This is incorrect. Because hospital mortality rates are assumed to be derived from a normal distribution with a fixed mean and variance, the random effects model will shrink small hospitals toward the only parameter it has regarding the hospital characteristic, that being the population mean. In fact, in practice, even large hospitals experience significant shrinkage of their predicted mortality rates toward the mean (see Figure 2, Silber et al. 2010, p. 1154). This result is due to the fundamental assumption in a random effects model, that any variable not included in the random effects model is assumed to be uninformative (Normand, Glickman, and Gatsonis 1997; Cameron and Trivedi 2005), and when specifically discussing hospital volume in an AMI mortality model (Normand and Shahian 2007). This is an important issue that leads to confusion when the strong implications of the model assumptions are underappreciated. As we have pointed out in our previous work, by not placing volume or other hospital characteristics in the HC model, HC assumes these variables are unimportant and produces posterior predictions that are shrunken to the mean of all hospitals. However, in our expanded model, using the same noninformative priors as in HC, and using the same dataset, the model predictions are very different. The expanded model also assumed that all other variables not included in the expanded model were unimportant, as this assumption is implied in the random effects model. We can judge whether it makes sense to include the expanded model hospital characteristics by asking which make better predictions. We show that the expanded model is by far a more accurate model than
the present HC model. However, we may find that still other models, which include other hospital variables and potentially interactions between hospital and patient characteristics, are better.

Hospital Compare should also address the question as to whether direct or indirect standardization should be utilized. HC uses indirect standardization, or the P/E approach. It compares how one hospital’s patients fared at that specific hospital versus how these same patients would perform at the typical hospital (defined as having the average hospital effect across all hospitals). HC multiplies P/E by the national mortality rate to provide simpler communication with the public. The problem with this approach is that indirect standardization can tell us whether a hospital is better than or worse than the national average, but it cannot formally tell us whether one hospital is better or worse than another, because such hospital comparisons are being made on a different mix of patients.

Instead, consider direct standardization. In our approach, we utilized the Bayesian model to compare the same mix of patients across all hospitals under the hypothetical that all patients went to all hospitals; that is, each hospital’s directly standardized mortality rate reflected the same exact patients (that being all the patients in the dataset), so a direct comparison can be made. A detailed explanation of the strengths and weaknesses of direct and indirect standardization has recently been published by Silber et al. (2014a,b), and standardization using Bayesian models is discussed in detail in George et al. (2015).

In conclusion, we have described potential enhancements to the HC model that may aid in helping consumers choose hospitals. First and foremost, Medicare should not exclude hospital characteristics from the HC model, and instead allow for the addition of important hospital characteristics that help the model make better predictions for patients seeking the best hospital. Second, the reporting of the improved HC model should consider using direct standardization when choosing between individual hospitals. Finally, we believe that in one form or another, the full posterior probability density function needs to be conveyed to the public. Whether this presentation is ultimately with density plots, bar graphs, or even video clips of this function, the information is useful. While future research is needed to perfect the ideal method of displaying Bayesian predictions to the public, it is now clear that HC should make these important modeling improvements to aid the public in making critically important decisions concerning hospital choice.
ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: We thank Traci Frank for her support in conducting this research. This work was funded by a grant from the Agency for Healthcare Research and Quality R21-HS021854; and grants SBS-1260782 and DMS-1406563 from the National Science Foundation.

Disclosures: None.

Disclaimers: None.

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article:

- Appendix SA1: Author Matrix.
- Data S1. Statistical Appendix.