Composite Quality Measures for Common Inpatient Medical Conditions

Lena M. Chen, MD, MS,†‡ Douglas O. Staiger, PhD,§ John D. Birkmeyer, MD,‡‖ Andrew M. Ryan, PhD, MA,¶ Wenying Zhang, MA,‡ and Justin B. Dimick, MD, MPH‡‖*

Background: Public reporting on quality aims to help patients select better hospitals. However, individual quality measures are suboptimal in identifying superior and inferior hospitals based on outcome performance.

Objective: To combine structure, process, and outcome measures into an empirically derived composite quality measure for heart failure (HF), acute myocardial infarction (AMI), and pneumonia (PNA). To assess how well the composite measure predicts future high and low performers, and explains variance in future hospital mortality.

Research Design: Using national Medicare data, we created a cohort of older patients treated at an acute care hospital for HF (n = 1,203,595), AMI (n = 625,595), or PNA (n = 1,234,299). We ranked hospitals on the basis of their July 2005 to June 2008 performance on the composite. We then estimated the odds of future (July to December 2009) 30-day, risk-adjusted mortality at the worst versus best quintile of hospitals. We repeated this analysis using 2005–2008 performance on existing quality indicators, including mortality.

Results: The composite (vs. Hospital Compare) explained 68% (vs. 39%) of variation in future AMI mortality rates. In 2009, if an AMI patient had chosen a hospital in the worst versus best quintile of performance using 2005–2008 composite (vs. Hospital Compare) rankings, he or she would have had 1.61 (vs. 1.39) times the odds of dying in 30 days (P-value for difference <0.001). Results were similar for HF and PNA.

Conclusions: Composite measures of quality for HF, AMI, and PNA performed better than existing measures at explaining variation in future mortality and predicting future high and low performers.

Key Words: hospital quality, medical, quality measurement

(*Med Care* 2013;51: 832–837)

From the *Department of Internal Medicine, Division of General Medicine, University of Michigan; †VA Health Services Research and Development, Center for Clinical Management Research, VA Ann Arbor Healthcare System; ‡Center for Healthcare Outcomes & Policy, University of Michigan, Ann Arbor, MI; §Department of Economics, Dartmouth University, Hanover, NH; ‖Department of Surgery, University of Michigan, Ann Arbor, MI; and ¶Weill Cornell Medical College, New York, NY.

Supported by the Agency for Healthcare Research and Quality, and National Institute on Aging.

J.D.B., J.B.D., and D.O.S. report having equity interest in ArborMetrix, Inc., a company that provides software and services for profiling hospital quality and costs. The company played no role in the manuscript. The remaining authors declare no conflict of interest.

Reprints: Lena M. Chen, MD, MS, Department of Internal Medicine, Division of General Medicine, University of Michigan, North Campus Research Complex, 2800 Plymouth Road, Building 16, Room 407E, Ann Arbor, MI 48109-2800. E-mail: lenac@umich.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

ISSN: 0025-7079/13/5109-0832

The Centers for Medicare and Medicaid Services (CMS) have launched numerous initiatives aimed at improving the quality of inpatient care, including public reporting on hospital performance on websites such as Hospital Compare. Proponents of public reporting hope that it will improve quality by motivating providers to engage in quality improvement and by guiding patients to high-quality hospitals.[1] For the latter mechanism to be successful, it would be important to have quality measures that reliably distinguish between the best and worst hospitals when patients use public reports.

However, there is growing recognition that existing quality measures are suboptimal in identifying superior and inferior hospitals on outcome performance. In 2011, the Medicare Payment Advisory Commission convened a technical panel to discuss weaknesses of current quality measures. Structural measures such as volume are poor proxies for outcome, especially at high-volume hospitals where most medical patients receive their care.[2] Performance on process measures is weakly associated with mortality for common medical conditions.[3] Furthermore, outcome measures such as risk-standardized mortality rates calculated using CMS' model do not account for the association between smaller volume and worse outcomes in patients with acute myocardial infarction (AMI).[4] Finally, lack of parsimony in publicly reported quality measures may provide conflicting guidance.

Composite measures of quality may offer better guidance to payers, patients, and providers seeking to distinguish high-quality hospitals from low-quality hospitals on outcome performance.[5] Empirically derived composites combine individual quality metrics—such as structure, process, and outcome indicators—into a single measure, weighting each input measure based on its reliability and correlation with the outcome of interest for a given condition (eg, AMI mortality). For surgical conditions, prior research has demonstrated that composite measures are better at predicting future mortality than volume or mortality alone.[6] A similar approach has not yet been applied to medical conditions. Therefore, using mortality as our gold-standard quality measure, we sought to create a composite quality measure for each of the 3 common medical conditions [heart failure (HF), AMI, and pneumonia (PNA)] that would predict future mortality. We then sought to assess the composite measure's performance relative to existing quality measures.

METHODS

Overview

For each condition studied, we combined individual structural, process, and outcome measures into a condition-specific composite measure of quality. The amount of hospital-level variance in 30-day, risk-adjusted mortality (RAM) that was explained by each measure determined both which measures to include in the composite and the weight to place on each measure included. We assessed the performance of the composite by evaluating the proportion of future hospital-level variation in mortality explained by the composite versus existing measures of quality, and the ability of each of these measures to discriminate between future high and low performers with respect to RAM.

Data Sources

We used data from the Medicare Provider Analysis and Review (MedPAR) files for 2005–2009. These files contain claims data for all acute care and critical access hospitalizations for fee-for-service Medicare beneficiaries. We obtained additional information on patient sociodemographics from the Medicare Denominator File for 2004–2008 and the US Census. Together, these data were used to calculate hospitals' 30-day RAM rates and condition-specific volume (a structural measure assessed for inclusion in the composite). We additionally obtained information from the American Hospital Association Annual Survey on several other structural measures assessed for inclusion in the composite, and from Hospital Compare for individual process measures assessed for inclusion in the composite.

Study Population

Our study sample included older Medicare beneficiaries who had been hospitalized with a principal diagnosis of HF (n = 1,203,595), AMI (n = 625,595), or PNA (n = 1,234,299) between July 1, 2005 and June 30, 2008. In order to compare the performance of the composite measure to that of existing hospital performance measures (eg, mortality as reported on Hospital Compare), we used similar inclusion and exclusion criteria as those used to create the cohort for Hospital Compare mortality measures (see Supplemental Digital Content 1, http://links.lww.com/MLR/A518 for additional details).[7]

Individual Quality Indicators Assessed for Inclusion in the Composite

To create the composite measure of quality, we examined the association between 30-day RAM and hospital performance on a broad range of individual quality measures. To estimate 30-day RAM, we first used a patient-level logistic regression model in which 30-day mortality was the dependent variable and covariates included demographics (eg, age, sex, and race), socioeconomic status,[8] urgency of admission (emergent/urgent), and comorbid conditions using methods defined by Elixhauser et al[9] [Healthcare Cost and Utilization Project (HCUP) Comorbidity Software, Version 3.3].[10] We obtained the predicted probability of the outcome for each patient and then summed these probabilities by hospital to estimate a hospital-specific expected mortality rate. Thirty-day RAM was then calculated by dividing the observed by the expected deaths, and multiplying this by the overall average condition-specific mortality rate.

We assessed 3 types of quality measures for inclusion in the composite measure: structural, process, and outcome measures. Structural measures included both volume indicators and other hospital characteristics. We evaluated 3-year (July 2005 to June 2008) hospital volume for HF, AMI, and PNA (only for Medicare beneficiaries in our cohort). We also evaluated volume for related conditions. On the basis of clinical judgment, related conditions for HF were defined as aortic valve repair (AVR), coronary artery bypass grafting (CABG), percutaneous coronary intervention (PCI), mitral valve repair (MVR), AMI, and PNA (see Supplemental Digital Content 1, http://links.lww.com/MLR/A518 for ICD-9 codes used). Related conditions for AMI were defined as AVR, CABG, PCI, MVR, and HF. The single related condition for PNA was HF. We also examined other structural measures such as teaching status, number of beds, presence or absence of an intensive care unit, proportion of Medicare days/total facility inpatient days, proportion of Medicaid days/total facility inpatient days, and hospital region.

We assessed several process measures (4 for HF, 7 for AMI, and 7 for PNA) for inclusion in the composite (see Supplemental Digital Content 2, http://links.lww.com/MLR/A519 for process of care measures). Hospital performance on each of these measures was recorded from July 2007 to June 2008, and reported on Hospital Compare in March of 2009. The outcome measures that we assessed for inclusion in the composite were mortality rates for HF, AMI, and PNA, as well as 30-day risk-standardized readmission rates for the same conditions.

Selection of Individual Measures for the Composite and Weighting of these Measures

From the structural, process, and outcome measures described above, we selected measures for inclusion in the composite using the following approach. Condition-specific mortality and volume were always included in the model. Similarly to prior work on surgical composite measures,[11] quality indicators that explained >10% of variation in hospital-level RAM were also included in the composite (Table 1). The composite measure was then calculated as the weighted sum of RAM and expected mortality (ie, expected

TABLE 1. Proportion of Hospital-Level Variation in Mortality Rates Explained by Individual Quality Measures Included in the Composite

Medical Condition	Individual Quality Measures	Proportion of Hospital-Level Variation Explained (%)
HF	Mortality	60
	PNA mortality	42
	Volume	23
	AMI mortality	21
	Related volume*	17
	No. beds	12
AMI	Mortality	56
	Volume	42
	Related volume†	41
	HF mortality	22
	PNA mortality	22
	No. beds	17
	RN hours per patient day	15
	Aspirin on discharge	12
	Aspirin on arrival	11
PNA	Mortality	70
	HF mortality	36
	AMI mortality	17
	Related volume‡	7
	Volume	4

The proportion of hospital-level variation explained does not sum to 100% because the variation explained by each component is not unique.

*For HF, related conditions were coronary artery bypass grafting (CABG), aortic valve repair (AVR), percutaneous coronary interventions (PCI), mitral valve repair (MVR), AMI, and PNA.

†For AMI, related conditions were AVR, CABG, PCI, MVR, and HF.

‡For PNA, the related condition was HF.

AMI indicates acute myocardial infarction; HF, heart failure; PNA, pneumonia; RN, registered nurse.

TABLE 2. Average Weights Across Hospitals Given to Input Measures Included in the Composite

Medical Conditions	Individual Quality Measures	Weight in the Composite Measure (%)
HF	Mortality	44
	Mortality expected given volume and other structural factors	29
	PNA mortality	17
	AMI mortality	10
AMI	Mortality	43
	Mortality expected given structural and process factors	37
	HF mortality	13
	PNA mortality	7
PNA	Mortality	57
	Mortality expected given volume	22
	HF mortality	16
	AMI mortality	5

AMI indicates acute myocardial infarction; HF, heart failure; PNA, pneumonia.

mortality given individual quality indicators included in the composite as well as patient risk factors). This methodology has been described in detail in earlier work,[12] and we expand on our methodology in Supplemental Digital Content 3, http://links.lww.com/MLR/A520. The average weights across hospitals were based on the amount of additional hospital-level variation explained by each measure, after all measures selected for inclusion in the composite were included in the model (Table 2).

Validation of the Composite

We determined the value of our composite measure in 3 ways. First, we compared the proportions of hospital-level variance in future 30-day mortality explained by the composite measure versus existing quality measures, after adjusting for patient covariates (see Supplemental Digital Content 3, http://links.lww.com/MLR/A520 for details). Because patients, payers, and providers typically make inferences about current hospital performance based on historical reports, we validated our composite measure in 2 additional ways. We first estimated how well historical (July 2005 to June 2008) hospital rankings, using the composite measure, predicted variance in future (June 2009 to December 2009) RAM. We also compared how well historical hospital rankings based on the composite versus other quality indicators discriminated between future low and high performers with respect to RAM.

For this latter step, we first ranked hospitals based on their historical performance on 5 measures of quality (ie, first volume, then RAM, then an aggregate of process measures using a methodology previously defined,[13] then mortality as reported on Hospital Compare, and finally the composite measure). For each of the 5 measures of quality, we then classified each hospital into a performance quintile of equal patient size (based on admissions from July 2005 to June 2008). For each of the 5 measures of quality, we then calculated the odds of future (July to December 2009) mortality at a hospital in the worst quintile of performance (1-star hospitals) versus the best quintile of performance (3-star hospitals). We were left with 5 separate odds ratios, 1 for each measure of quality. We used nonparametric bootstrapping with replacement to assess whether the odds ratio for the composite measure was statistically different (at the $P<0.05$ level) from odds ratios that relied on other quality measures to rank hospitals.

We conducted all analyses using SAS version 9.2 and Stata 11.2. The study protocol was approved by the Institutional Review Board at the University of Michigan.

RESULTS

Components of the Composite Measure

Among individual quality measures, mortality for the condition of interest explained the largest proportion of hospital-level variance in RAM for that same condition (Table 1), where the hospital-level variance comes from the variance-covariance matrix of the hospital-level quality parameters [see equation (3) in Supplemental Digital Content 3, http://links.lww.com/MLR/A520]. For example, HF, AMI, and PNA mortality, respectively, explained 60%, 56%, and 70% of hospital-level variance in HF, AMI, and PNA RAM. Performance for HF and PNA were related, as PNA mortality explained 42% of hospital-level variance in HF RAM. This

TABLE 3. Proportion of Future Hospital-Level Variance Explained by Existing Quality Measures Versus the Composite Measure

	Proportion of Variance Explained (%)				
Conditions	Hospital Volume	Hospital Compare Process Measures	Risk-adjusted Mortality	Hospital Compare Outcome Measures	Composite Measure
HF	8	0	37	32	44
AMI	35	6	47	39	68
PNA	4	1	32	24	38

AMI indicates acute myocardial infarction; HF, heart failure; PNA, pneumonia.

was the second largest explanatory factor for HF RAM. Similarly, HF mortality explained 36% of hospital-level variance in PNA RAM. This was the second largest explanatory factor for PNA RAM.

Among structural factors, volume explained the largest proportion of hospital-level variation and was most important for AMI. For example, AMI volume explained 42% of hospital-level variation in AMI RAM and was the second largest explanatory factor. Related volume (ie, volume for the related conditions of AVR, CABG, PCI, MVR, and HF) explained a similar proportion of hospital-level variation in AMI RAM. In contrast, volume and related volume were not as important for HF RAM, explaining 23% and 17%, respectively, of hospital-level variation in HF RAM. Volume and related volume explained only 4% and 7%, respectively, of hospital-level variation in PNA RAM. A few other structural factors such as number of beds explained a small proportion of hospital-level variation in RAM for HF and AMI.

None of the individual process measures explained a large enough proportion of hospital-level variation in RAM (ie, >10%) to be included in the composite for HF or PNA. Two process measures were included in the composite for AMI. Aspirin on discharge and aspirin on arrival, respectively, explained 12% and 11% of hospital-level variation in RAM for AMI. Of note, in sensitivity analyses, we restricted the patient sample to the time frame of the process measures (ie, July 2007 to June 2008) and obtained similar results.

In the final composite measures, the largest weight was placed on mortality for the condition of interest (Table 2). For example, HF, AMI, and PNA mortality, respectively, received 44%, 43%, and 57% of the weight for the HF, AMI, and PNA composite measures. Mortality expected given structural measures and (for AMI) process measures received the second largest weight in the composite measures.

Patient and Hospital Characteristics by Quintiles of Performance on the Composite

Hospitals in different quintiles of performance on the composite measure differed in the types of patients treated and in structural characteristics (see Supplemental Digital Content 4, http://links.lww.com/MLR/A521 for descriptive statistics). For all 3 conditions, poorer patients were less likely to be cared for at 3-star (top quintile) versus 1-star (bottom quintile) hospitals. Consistent with prior work on the association between volume and outcomes, higher-performing hospitals were larger and had more experience treating patients with the condition of interest. High-performing hospitals were also more likely to be teaching hospitals and much less likely to be critical access hospitals.

Ability of the Composite to Predict Future Performance

Hospital rankings based on the composite explained a greater proportion of hospital-level variation in future mortality rates than hospital rankings based on Hospital Compare. Hospital performance on the composite explained 44%, 68%, and 38% of the variation in future HF, AMI, and PNA mortality rates, respectively. In contrast, hospital rankings based on Hospital Compare mortality rates explained 32%, 39% and 24% of the variation in future HF, AMI, and PNA mortality rates, respectively (Table 3).

For all 3 conditions, historical performance on the composite measure predicted future high and low performers with respect to RAM (Fig. 1). For example, AMI patients

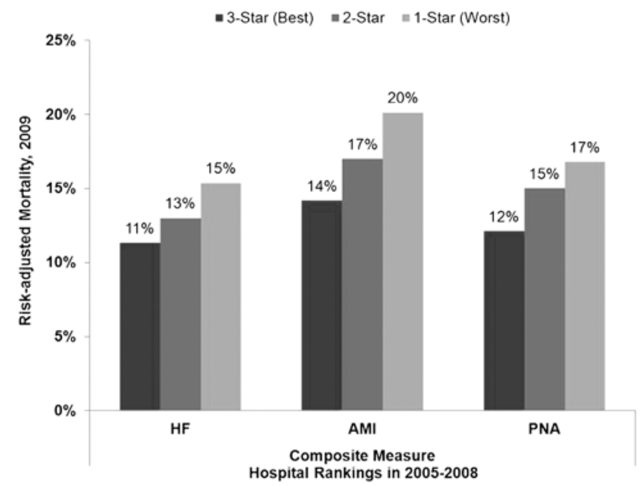

FIGURE 1. Future risk-adjusted mortality rates (July 2009 to December 2009) for 1-star, 2-star, and 3-star hospitals (ranked using the composite measure and July 2005 to June 2008 data). One-star hospitals were those hospitals in the worst quintile of performance when using the composite measure with July 2005 to June 2008 data. Three-star hospitals were those hospitals in the best quintile of performance when using the composite measure with July 2005 to June 2008 data. Two-star hospitals were all other hospitals. AMI indicates acute myocardial infarction; HF, heart failure; PNA, pneumonia.

TABLE 4. Relative Ability of Historical Hospital Rankings Based on Different Quality Measures, to Forecast Future Risk-adjusted Mortality, for All Hospitals

	Adjusted Odds Ratio for Risk-adjusted Mortality in 2009 (July to December), 1-Star (Bottom 20%) Versus 3-Star (Top 20%) Based on July 2005 to June 2008 Hospital Rankings (95% CI)				
Conditions	**Hospital Volume**	**Hospital Compare Process Measures**	**Risk-adjusted Mortality**	**Hospital Compare Mortality Measures**	**Composite Measure**
HF	1.19 (1.14–1.25)	1.01 (0.97–1.06)	1.42 (1.35–1.48)	1.38 (1.31–1.44)	1.47 (1.41-1.54)
AMI	1.38 (1.31–1.47)	1.10 (1.08–1.12)	1.45 (1.37–1.54)	1.39 (1.31–1.48)	1.61 (1.52–1.71)
PNA	1.17 (1.12–1.23)	1.06 (1.01–1.11)	1.48 (1.41–1.55)	1.42 (1.35–1.49)	1.54 (1.46–1.61)

P-value for difference (between odds ratio for composite measure and odds ratio for individual measures) is <0.05 for all comparisons.
AMI indicates acute myocardial infarction; CI, confidence interval; HF, heart failure; PNA, pneumonia.

treated at historically ranked 1-star hospitals (ie, in the bottom quintile of performance on the composite based on July 2005 to June 2008 data), had a 20% RAM rate in the second half of 2009. In contrast, AMI patients treated at historically ranked 3-star hospitals (ie, in the top quintile of performance on the composite based on July 2005 to June 2008 data), had a 14% RAM rate in the second half of 2009. The differences in future performance between historically ranked 1-star, 2-star, and 3-star hospitals were smaller for HF and PNA.

Historical performance on the composite measure was better able to discriminate between high-performing and low-performing hospitals (Table 4). For example, in the second half of 2009, if an AMI patient had chosen a hospital in the worst versus best quintile of performance using July 2005 to June 2008 composite (vs. Hospital Compare) rankings, he or she would have had a 61% (vs. 39%) greater odds of dying ($P<0.001$). Volume, RAM, and aggregate process measures also performed worse than the composite in predicting future RAM. In the second half of 2009, the odds of dying at a hospital in the worst versus best quintile of performance for AMI using 2005–2008 rankings based on volume, process measures, or RAM, were 1.38, 1.10, and 1.45, respectively (*P*-values for difference with the odds ratio for the composite all <0.05). In sensitivity analyses restricting the sample to small hospitals (ie, the lowest quartile of hospitals when ranked by condition-specific volume), the performance of the composite relative to existing quality measures was similar (see Supplemental Digital Content 5, http://links.lww.com/MLR/A522 for analysis of small hospitals).

DISCUSSION

Proponents of public reporting hope that it will help patients seek care at high-quality hospitals. To achieve this goal, it would be important to report on quality measures that provide clear and reliable guidance about hospital quality. We found that a composite quality measure empirically incorporating multiple structural, process, and outcome measures was modestly better at predicting variance in future RAM, and at discriminating between future low and high performers than many existing quality measures. Among the 3 medical conditions that we studied, the composite performed best for AMI.

The value of the medical composite should be assessed in comparison with existing quality measures. Process measures have been found to be correlated with mortality,[14] although the association is not strong,[3] and a correlation with longer-term outcomes has not always been identified.[15] Structural measures such as volume are associated with mortality for medical conditions.[16] However, for HF, AMI, and PNA, the association between volume and outcomes is strongest at low volumes and most patients are seen at hospitals with higher volumes.[2] Outcomes measures such as mortality have wide year-to-year variation, which permits past performance to reliably identify only extreme outliers. Shrunken estimates, such as those derived from the model used by Hospital Compare, can pull estimates for low-volume hospitals towards the average for all hospitals.[4]

Composite measures have been explored in medicine,[17–19] but one of the most widely used composites—that employed by CMS to combine performance on process measures for HF, AMI, and PNA—equally weights all inputs. Our empirically derived composite is distinct in that it draws on multiple inputs (ie, existing process, structural, and outcome measures), and weights each measure in order to best predict a concrete outcome: RAM. As such, it is easily reproducible, and can quickly adapt to changes in hospital performance along a number of dimensions. The composite may be revised overtime to include new measures and/or to incorporate new empirically estimated weights that may be changing.

To create the composite, we applied a method that has previously been used to construct composite measures of quality for surgical conditions.[6] In comparison with the predictive ability of surgical composites, the predictive ability of medical composites was more modest. For example, for surgical patients treated at historically ranked 1-star versus 3-star hospitals, the odds ratios for future RAM were 2.10 (CABG, AVR), 3.29 (pancreatic cancer resection), and 3.91 (esophageal cancer resection).[11] In comparison, for medical patients treated at historically ranked 1-star versus 3-star hospitals, the odds ratios for future RAM were 1.47 (HF), 1.61 (AMI), and 1.54 (PNA).

The weaker predictive ability of medical compared to surgical composites may be explained by at least 2 factors. First, even though volume is associated with outcomes for both medical and surgical conditions, volume has been found to have a stronger association with surgical mortality.[2,20]

In surgery, technical mastery depends considerably on practice. Both medical and surgical composites draw heavily on volume to predict future mortality. Thus, it is perhaps not surprising that the composite performed best for AMI, where outcomes are often related to performance on a procedure (PCI). Second, there is likely wider variation in case mix for medical compared to surgical patients. Surgery may be deferred if risk is too high, but all admitted medical patients are offered treatment. Because of this, the association between unmeasured heterogeneity in case mix and mortality is presumably greater for medical as compared with surgical patients.

Of note, we did not validate the composite by comparing predictions of individual RAM rates to actual future RAM rates. We used an out-of-sample prediction of a classification (star ratings) rather than an out-of-sample prediction of a rate. The star classification system is more consistent with how profiling is implemented in practice. However, prior work has used simulation (where the "true" underlying rates were known) to compare the composite measure to other approaches of measuring hospital quality with AMI. The composite performed well in comparison to other hospital quality outcome measures.[5]

Our study has several potential limitations. First, we studied 3 conditions in the Medicare population. However, these conditions are common and those most likely to be hospitalized for them are older patients. Moreover, we chose high-volume conditions that should be least likely to demonstrate the strengths of a composite measure that includes volume. Second, we chose to predict RAM. As described above, this is a strength. At the same time, this approach provides no information about other outcomes, such as functional capacity or overall satisfaction with care. When patient-level data on outcomes other than mortality are widely available, an empirically derived composite could be used to predict these outcomes. Finally, the practical utility of the composite depends on the public's comfort with a composite measure of quality, as opposed to the more intuitive measure of RAM. We did not have the ability to assess how patients or providers might interpret the composite.

In summary, public reporting aims to improve the quality of care delivered to patients. This goal depends on the use of quality measures that reliably identify high-quality and low-quality hospitals for patients and providers. Reliable measures would help patients identify the best hospitals for their care, and help providers better identify hospitals with best practices to be emulated. In this context, we found that composite measures of quality for HF, AMI, and PNA are modestly better than existing measures at explaining variation in future mortality and at predicting future high and low performers with respect to RAM.

REFERENCES

1. Fung CH, Lim YW, Mattke S, et al. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med*. 2008;148:111–123.
2. Ross JS, Normand SL, Wang Y, et al. Hospital volume and 30-day mortality for three common medical conditions. *N Engl J Med*. 2010; 362:1110–1118.
3. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006;296:2694–2702.
4. Silber JH, Rosenbaum PR, Brachet TJ, et al. The Hospital Compare mortality model and the volume-outcome relationship. *Health Serv Res*. 2010;45:1148–1167.
5. Ryan A, Burgess J, Strawderman R, et al. What is the best way to estimate hospital quality outcomes? A simulation approach. *Health Serv Res*. 2012;47:1699–1718.
6. Dimick JB, Staiger DO, Baser O, et al. Composite measures for predicting surgical mortality in the hospital. *Health Aff (Millwood)*. 2009;28:1189–1198.
7. QualityNet. Hospitals Inpatient-Claims-based Measures-Mortality Measures.
8. Birkmeyer NJ, Gu N, Baser O, et al. Socioeconomic status and surgical mortality in the elderly. *Med Care*. 2008;46:893–899.
9. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.
10. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). HCUP Comorbidity Software. Available at: http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp. Accessed July 11, 2013.
11. Dimick JB, Staiger DO, Osborne NH, et al. Composite measures for rating hospital quality with major surgery. *Health Serv Res*. 2012;47: 1861–1879.
12. Staiger DO, Dimick JB, Baser O, et al. Empirically derived composite measures of surgical performance. *Med Care*. 2009;47:226–233.
13. Kahn CN III, Ault T, Isenstein H, et al. Snapshot of hospital quality reporting and pay-for-performance under Medicare. *Health Aff (Millwood)*. 2006;25:148–162.
14. Jha AK, Orav EJ, Li Z, et al. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood)*. 2007;26:1104–1110.
15. Patterson ME, Hernandez AF, Hammill BG, et al. Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. *Med Care*. 2010;48:210–216.
16. Joynt KE, Orav EJ, Jha AK. The association between hospital volume and processes, outcomes, and costs of care for congestive heart failure. *Ann Intern Med*. 2011;154:94–102.
17. Peterson ED, DeLong ER, Masoudi FA, et al. ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment: a report of American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (writing committee to develop a position statement on composite measures). *J Am Coll Cardiol*. 2010;55:1755–1766.
18. Reeves D, Campbell SM, Adams J, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care*. 2007;45:489–496.
19. Shwartz M, Ren J, Pekoz EA, et al. Estimating a composite measure of hospital quality from the Hospital Compare database: differences when using a Bayesian hierarchical latent variable model versus denominator-based weights. *Med Care*. 2008;46:778–785.
20. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346: 1128–1137.

