U.S. Department of Veterans Affairs
Public Access Author manuscript
Healthc (Amst). Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:
Healthc (Amst). 2017 September ; 5(3): 112–118. doi:10.1016/j.hjdsi.2016.10.001.

Does adding clinical data to administrative data improve agreement among hospital quality measures?

Amresh D. Hanchate, PhD[1,2], **Kelly L. Stolzmann, MS**[1], **Amy K. Rosen, PhD**[1,3], **Aaron S. Fink, MD**[4], **Michael Shwartz, PhD**[1,5], **Arlene S. Ash, PhD**[6], **Hassen Abdulkerim, MS**[1], **Mary Jo V. Pugh, PhD**[7,8], **Priti Shokeen, MS**[1], and **Ann Borzecki, MD MPH**[2,9,10]

[1]Center for Healthcare Organization and Implementation Research (CHOIR), VA Boston Healthcare System, Boston, MA 02130

[2]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02118

[3]Department of Surgery, Boston University School of Medicine, Boston, MA 02118

[4]Professor Emeritus of Surgery, Emory University School of Medicine, Atlanta, GA 30322

[5]Department of Operations and Technology Management, Boston University School of Management, Boston, MA 02215

[6]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA 01655

[7]South Texas Veterans Health Care System, San Antonio, TX 78229

[8]Department of Epidemiology and Biostatistics, University of Texas Health Science Center, San Antonio, TX 78229

[9]Center for Healthcare Organization and Implementation Research (CHOIR), Bedford VAMC, Bedford, MA 01730

[10]Boston University School of Public Health, Boston, MA 02118

Abstract

Background—Hospital performance measures based on patient mortality and readmission have indicated modest rates of agreement. We examined if combining clinical data on laboratory tests and vital signs with administrative data leads to improved agreement with each other, and with other measures of hospital performance in the nation's largest integrated health care system.

Methods—We used patient-level administrative and clinical data, and hospital-level data on quality indicators, for 2007-2010 from the Veterans Health Administration (VA). For patients admitted for acute myocardial infarction (AMI), heart failure (HF) and pneumonia we examined changes in hospital performance on 30-day mortality and 30-day readmission rates as a result of

Corresponding Author and Requests for Reprints: Amresh D. Hanchate, PhD, Associate Professor, Health/care Disparities Research Program, Section of General Internal Medicine, Boston University School of Medicine, 801 Massachusetts Ave, #2092, Boston, MA 02118, Tel: (617) 638-8889, Fax: (617) 638-8026, hanchate@bu.edu.

Conflicts of Interest: None of the authors have conflicts or potential conflicts of interest. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or Boston University.

adding clinical data to administrative data. We evaluated whether this enhancement yielded improved measures of hospital quality, based on concordance with other hospital quality indicators.

Results—For 30-day mortality, data enhancement improved model performance, and significantly changed hospital performance profiles; for 30-day readmission, the impact was modest. Concordance between enhanced measures of both outcomes, and with other hospital quality measures – including Joint Commission process measures, VA Surgical Quality Improvement Program (VASQIP) mortality and morbidity, and case volume – remained poor.

Conclusions—Adding laboratory tests and vital signs to measure hospital performance on mortality and readmission did not improve the poor rates of agreement across hospital quality indicators in the VA.

Interpretation—Efforts to improve risk adjustment models should continue; however, evidence of validation should precede their use as reliable measures of quality.

Keywords

clinical data; hospital quality; 30-day mortality; 30-day readmission; Hospital Compare

With growing momentum for greater transparency and accountability of gaps in hospital quality, the range of measures of hospital quality has steadily grown, calling for a better understanding of the level of agreement among them.[1–3] Of particular significance are the Centers for Medicare and Medicaid Services' (CMS) Hospital Compare measures, given their conspicuous profile in the quality measurement landscape, and their instrumental role as the basis for determining rewards and penalties for CMS' Value-Based Purchasing and Hospital Readmissions Reduction programs.[2, 4, 5] Recent studies have evaluated agreement among Hospital Compare measures and other quality indicators, on the premise that these measures together "reflect a construct of core hospital quality" and that "a hospital deemed high quality would perform well across a variety of domains of care".[6] The overall consensus in findings indicates poor agreement among quality indicators.[7] One study compared Hospital Compare rates of 30-day mortality with 30-day readmission for patients admitted for acute myocardial infarction (AMI), heart failure (HF) and pneumonia, and found weak to no correlation for all cohorts.[8] Other studies compared performance on mortality with that on compliance with process of care measures and generally found poor agreement for several medical and surgical admissions.[9–12] Patient volume, a structural indicator widely associated with outcome quality, was also found to be weakly correlated with readmission rates.[6]

Given the central focus on patient outcome measures in the aforementioned comparative studies, a possible explanation for poor concordance is the limited clinical content in the administrative data used to account for differences in patient health status at admission. Skepticism over the use of administrative data to measure hospital quality dates back to the origin of report cards nearly two decades ago, with particular emphasis on the limitations of diagnostic and procedure codes to adequately capture patient severity at admission.[5, 13, 14] To address this limitation, several initiatives have supplemented administrative data with clinical measures of patient status at or near admission in order to evaluate hospital

performance. One promising avenue of enhanced risk adjustment, currently being evaluated in pilot settings by the Agency for Healthcare Research and Quality (AHRQ) and other stakeholders, is the addition of data on laboratory tests and vital signs for evaluation of patient mortality.[15, 16] Several studies, based on convenience samples of hospitals, have found that adding data on laboratory tests and vital signs, measured at the time of admission, significantly improved the ability of models to discriminate patient risk for mortality and readmission.[17–19]

Our aim in this study was to examine whether adding laboratory tests and vital signs to obtain risk adjusted rates of mortality and readmission would lead to improved agreement among hospital quality measures. We used the setting of the Veterans Health Administration (VA), the nation's largest integrated health care system with 152 hospitals serving 8.5 million enrollees.[20] VA's integrated health care information system has been used extensively for quality assessment and reporting, as part of ongoing national programs and through unique in-house initiatives.[20–22] We modified the Hospital Compare 30-day mortality and 30-day readmission performance measures by adding data on laboratory tests and vital signs, and a) measured the impact on hospital performance indicators, and b) evaluated the concordance between the enhanced outcome measures, and with other hospital performance measures reflecting inpatient processes of care and hospital structures.

METHODS

The study involved two phases: in the first, we developed mortality and readmission performance measures using enhanced data; in the second, we evaluated the agreement between enhanced mortality and readmission measures, and between enhanced performance measures and other hospital quality measures. This study was approved by the VA Boston Healthcare System Institutional Review Board.

Data Sources

We used VA patient databases covering inpatient stays, outpatient visits, laboratory tests, vital signs and vital status (2006-2010).[23] These cover services provided at all VA hospitals and outpatient clinics, and include results of laboratory tests and vital signs performed in inpatient and outpatient settings.

Study Cohorts and Risk Measures

Using only administrative data, we applied the CMS Hospital Compare protocol ("administrative data model") to obtain risk adjusted hospital-level rates of 30-day mortality and 30-day readmission separately for the three admission cohorts;[21, 24] the only difference was that in our models, all patients aged 18 or older were included, whereas the Hospital Compare program includes only those 65 and older. Using VA acute inpatient discharge data for fiscal years 2007-2010, we identified all admissions, henceforth termed "index admissions", for patients with a principal diagnosis of AMI, HF and pneumonia using the International Classification of Diseases (ICD-9-CM) codes and exclusion criteria used by the Hospital Compare program.[25, 26]

In adding clinical data we identified risk measures from results of laboratory tests and vital signs performed within 24 hours, before or after, the time of the index admission; these included tests performed in outpatient care settings. We examined alternative time windows and found that (a) approximately 40% of tests were only identified in the 24 hours after admission time, (b) extending the time window beyond 24 hours did not increase the number of tests captured (Appendix A). Development of these enhanced measures was a multistep process and has varied across previous studies.[17–19, 27, 28] The steps we used, detailed in the supplementary materials (Appendix A), reflect the most common of the approaches used in the literature. Based on prior studies, clinical guidance on tests typically performed on most patients admitted for the selected conditions and completeness of data on test results across patients, we selected 16 laboratory tests (hemoglobin, potassium, sodium, blood urea nitrogen (BUN), white blood cell count (WBC), aspartate amino transferase (AST), glucose, creatinine, bilirubin, alkaline phosphatase, albumin, hematocrit, prothrombin time, partial prothrombin time, troponin and carbon dioxide/HCO3) and 6 vital signs (pulse, pulse oximetry, respiration, temperature and blood pressure [systolic and diastolic]) for which data were available for a majority of patients. Using a range of test values informed by clinical judgement, we performed bivariate correlations between mortality and the test values and categorized each test result into a maximum of 5 categories: normal, low abnormal, moderate abnormal, high abnormal and missing. Normal category refers to the range of test values with the lowest risk of mortality in bivariate analysis; abnormal categories indicate other test value ranges with higher risk of morality (Appendix A). We treated patients with a missing laboratory test result as a separate category so as to capture the risk associated with the decision not to perform the test; we also looked for systematic differences rates of missing test results across hospitals and time (Appendix A). In cases with multiple tests within 24 hours of admission, following prior work, we selected the most abnormal test reading.[17–19, 22] We excluded clinically implausible test results (Appendix A). For comparison and as a sensitivity test we examined an alternative categorization of laboratory tests and vital signs using thresholds commonly used in routine clinical practice (Appendix B). Based on preliminary logistic regression models we selected the final subset of laboratory tests and vital signs added to the measures from the administrative data model for each outcome and cohort (“enhanced data model”). All the analyses – categorization of test values and enhanced data model estimates – were not sensitive to use of out-of-sample data; we have reported estimates based on using combined data for better precision of estimates.

Risk Adjusted Mortality Rates

Using the administrative and enhanced data models, we followed the Hospital Compare protocol and obtained hospital-level risk adjusted mortality rates (RSMR) and readmission rates (RSRR) based on estimates of logistic and hierarchical logistic regression models. We estimated the 95% confidence intervals corresponding to the RSMR and RSRR estimates using bootstrap samples (N=1,000).[21] Hospital performance was grouped into three categories based on whether the confidence interval was entirely above (“worse than average”), or entirely below (“better than average”), or included (“no different from average”) the VA national mortality rate. This categorization of performance differs from that used by the Hospital Readmissions Reduction Program, wherein a hospital is designated as having “excess readmission ratio” if hospital RSRR exceeds the national readmission rate

for any one of the three admission cohorts.[29] We also grouped performance by quintiles of RSMR and RSRR. We compared performance of the administrative and enhanced data risk adjustment models using a range of indicators: discrimination (c-statistic), calibration (ratio of observed mortality between highest and lowest deciles of predicted mortality) and pseudo-R^2.[30] Using the bootstrap method we estimated the 95% confidence interval of these statistics using 500 bootstrap samples.

Impact of enhanced data on hospital performance

Adding clinical data can cause the predicted probability of the outcome to increase in some patients (and hospitals) and decrease in others; this is because the overall sum of individual probabilities equals the observed outcome rate, which is unchanged. Accordingly, we calculated the impact of using enhanced data on RSMR and RSRR in terms of absolute (% change) and relative (Hospital Compare performance designation and quintile) change.

Other Hospital Quality Indicators

In addition to RSMR and RSRR, we identified hospital quality indicators based on prior studies.[6, 8–10] Some measures (process of care) are produced as part of ongoing national programs, while others (surgical mortality and surgical morbidity) were introduced as VA initiatives.

Process of care measures—We obtained composite performance scores on the Joint Commission ORYX process measures of inpatient quality for AMI (7 measures), HF (4 measures) and pneumonia (6 measures[20, 31]) between 10/1/2008 and 9/30/2009.

Surgical mortality and morbidity measures—Using chart abstracted data for patients receiving a wide range of inpatient surgical procedures, and estimates of validated models of risk adjustment, VA Surgery Quality Improvement Program (VASQIP) provides hospital level ratios of observed to expected rates of 30-day mortality and morbidity.[32, 33] We used hospital ratios for mortality and morbidity for patients who received surgeries between 10/1/2008 and 9/30/2009.[20]

Case volume—For each admission cohort, this was defined as the number of index admissions in each hospital during the study period (2007-2010).

Concordance among Hospital Performance Indicators

We estimated concordance (kappa statistic) between quintiles of RSMR and RSRR before and after enhancement using clinical data. Similarly, we estimated concordance between RSMR or RSRR and other hospital performance indicators. In addition to the kappa statistic, we also estimated correlation and rank correlation using continuous performance measures; due to similar findings, only the concordance estimates (kappa statistics) are reported.

Temporal stability of RSMR and RSRR

To examine temporal stability of RSMR and RSRR, we divided the four-year study period into two 2-year periods (2007-2008 and 2009-2010) and compared hospital performance between the two periods (separately for administrative and enhanced data models).[6]

Regression to the mean

As RSMR and RSRR are statistical estimates, the change in each from adding enhanced data could partly be due to regression to the mean.[34] This occurs because hospitals with higher (lower) than the expected rate by one method (administrative data) are more likely to experience a decrease (increase) in the rate using the other method (enhanced data).[35, 36] To test for this phenomenon, we estimated a linear regression of the *change in adjusted rate* on the *adjusted rate prior to data enhancement*, and measured the variation arising from regression to the mean (r-squared). We then adjusted for the expected change from regression to the mean and re-estimated the linear regression.[34]

All analyses were performed using SAS 9.2 and Stata 13.1.

RESULTS

Nationally, the overall number of VA hospitals, by cohort, was 91 (AMI), 128 (HF) and 131 (pneumonia) for examining 30-day mortality and 97 (AMI), 130 (HF) and 131 (pneumonia) for examining 30-day readmission (Table 1). Average observed 30-day mortality/ readmission was 9.7% / 20.1% (AMI), 7.9% / 22.3% (HF), and 10.2% / 16.4% (pneumonia), and varied considerably across hospitals; for instance, average AMI mortality across hospitals ranged from 3.0% to 26.2%.

Impact of Adding Clinical Data on Risk Adjustment Model Performance

Table 2 indicates the change in model performance after adding clinical data; detailed estimates from the hierarchical logistic regression models are presented in Appendix A. In general, model performance improved substantially for all three mortality cohorts, but only modestly for the readmission cohorts. For mortality, model discrimination (c-statistic) improved from 0.79 to 0.85 (AMI), 0.73 to 0.81 (HF), and, 0.76 to 0.82 (pneumonia). Replication of the analyses using as thresholds for the normal range of laboratory tests and vital signs those used in routine clinical practice indicated similar findings, although the improvement in model discrimination was marginally smaller (Appendix B).

Impact of Adding Clinical Data on Hospital Performance

Measured in multiple ways, adding clinical data resulted in substantial changes in mortality performance but little change in readmission performance. Grouped into RSMR quintiles, we found that a large proportion of hospitals – 51% (AMI), 48% (HF) and 50% (pneumonia) – experienced change in the quintile group following the addition of clinical data (Table 3). Allowing for shifts across adjacent quintiles, we find many hospitals experienced shifts across 2 or more quintiles; 2 hospitals categorized in the lowest adjusted AMI mortality quintile were classified in 3rd and 4th quintile after enhanced risk adjustment, while 3 hospitals experienced a reverse change from the 4th quintile to the 1st or 2nd quintile. Hospital Compare performance designation also changed substantially: more hospitals were classified as not different from the VA national rate for AMI (85 to 90) and pneumonia (110 to 122), but for the HF cohort, fewer hospitals were classified as such (109 to 105) (Appendix C, Table C1). A closer examination of designation changes indicated that in roughly half the hospitals, this was accompanied by a sizable change (20% or more) in the

RSMR (Appendix C, Figure C1). Absolute RSMR changed 10% or more in over one out of five hospitals for all cohorts (Appendix C, Table C2). In contrast, absolute and relative RSRR experienced only modest changes.

Concordance between Mortality and Readmission Performance

Concordance between RSMR and RSRR, matched by cohort, was generally poor using administrative data across all three cohorts and remained poor after adding clinical data (Table 4).

Concordance of RSMR and RSRR with Other Quality Indicators

Using administrative data models, concordance of RSMR and RSRR with ORYX process scores and VASQIP mortality was poor, with a kappa statistic that was not different from 0 for all three cohorts (Table 5). Addition of clinical data did not change the concordance. Similarly, concordance with VASQIP morbidity, and case volume were also poor in both data settings (Appendix D).

As a measure of stability of mortality performance, we also compared RSMR during 2007-08 and 2009-10, and found that concordance was poor using both data models (Appendix D). Concordance between RSMR for pairs of different admission conditions showed no improvement after data enhancement (Appendix D). In both data settings, concordance for AMI versus HF and AMI versus pneumonia were not significantly different from 0. There was significant concordance for the HF versus pneumonia comparison using administrative data; however, following the addition of clinical data, concordance did not improve.

Change in RSMR and RSRR after Adding Clinical Data: Regression to the Mean

As an indication of the extent to which regression to the mean contributes to the change in RSMR and RSRR between administrative and enhanced data models, Figure E1 (Appendix) shows the correlation between *change in RSMR* and *RSMR prior to data enhancement*. For the AMI cohort, the correlation is significant and accounts for 53% of the variation in RSMR change across hospitals. Consistent with the regression to the mean phenomenon, hospitals with lower (higher) pre-enhancement RSMR experienced a larger (smaller) increase from data enhancement. We found a similar pattern for the pneumonia cohort, but found no correlation for the HF cohort. After adjusting for regression to the mean, we found no significant correlation for all three cohorts (Appendix E).

DISCUSSION

Adding clinical measures from laboratory tests and vital signs improved the performance of risk adjustment models for patient outcomes, particularly 30-day mortality. We found little evidence that this enhancement improves agreement among different indicators of hospital quality. The poor rates of agreement between hospital performance on mortality and readmission did not improve with the enhancement using clinical data. Agreement with other indicators of hospital quality based on process measures, surgical outcomes, and patient volume, also remained poor.

Our finding of improved model performance for 30-day mortality, using model discrimination as the criterion, is consistent with previous studies.[18, 19, 22] Even in relative terms, we found that many hospitals were reclassified from the lowest to the 4th quintile, or vice versa, following data enhancement. Our finding of modest improvement in discrimination for 30-day readmission is consistent with previous studies.[28]

How should we interpret the lack of improvement in agreement among hospital performance measures? First, improvement in model performance may not necessarily lead to more accurate measures, as often interpreted.[17–19, 22] Based on simulation analyses, Austin and Reeves report that improved c-statistic may not lead to improved accuracy in hospital profiling if the clinical data added are not "prognostically important" variables or if variation in these variables across hospitals is limited.[37] On both scores, evidence from VA seems to favor the enhanced risk adjustment model. First, clinical evidence points to increased mortality risk from abnormal laboratory tests;[38–40] and second, our data indicates considerable variation across hospitals in the prevalence of abnormal laboratory tests and vital signs. What is unclear is the extent of model performance improvement needed to elicit noticeable improvement in measurement accuracy.

Presence of statistical noise in the performance measures is another source of poor agreement across the measures. Although a quarter of VA hospitals experienced a change in risk adjusted mortality of 10% or more, regression to the mean was a prominent source of the change, contributing to 53% of the change in AMI mortality.

Lack of improvement in hospital performance indicators may also be due to shrinkage of estimates, particularly for low volume hospitals.[41] Austin and Reeves' simulation study found that higher hospital volume contributed more to accurate quality measurement than improved model performance.[37] Poor agreement could also be the result of competing risks for mortality and readmission; i.e., hospitals with high mortality have fewer patients at risk of readmission.[42, 43]

An alternative explanation for the poor agreement across performance measures is that the measures capture distinct dimensions of quality.[44] While process measures address specific elements of clinical care, mortality differences may be influenced by a wider range of treatment elements (early triage and care co-ordination) that may be correlated with structural differences (staffing and teaching hospitals).[45] Readmissions may be more sensitive to processes relating to discharge planning and follow-up care, as well as non-clinical factors (social supports and outpatient care access).[8, 46] These apparent distinctions are largely conjectures, with little formal research aimed at understanding the interrelationships between quality measures in concept and practice.

Our findings also speak to the increasing use of hospital quality measures for determining hospital rewards and penalties, although no such programs currently exist in the VA.[5] Our findings indicate that such programs may result in potentially puzzling pattern of payments: simultaneous bonuses and penalties for a sizable number of hospitals, and for the same quality measure, hospitals may alternate between bonuses and penalties from one year to the next.[47, 48] In the absence of clear process of care interventions proven to lead to

improvements in hospital quality indicators, there is concern that such incentives may lead to gaming behaviors.[3, 5, 6, 12, 49]

Our study has several limitations. In comparing several indicators of hospital quality with risk adjusted mortality and readmission rates we recognize that there is no gold standard measure; instead our rationale is that these measures have overlapping quality constructs.[6] As a large proportion of Veterans also receive care from non-VA providers, our characterization of patient comorbidities from VA data sources may be incomplete; however, previous work based on combining VA and Medicare data indicated only modest changes in patient risk.[50]

To summarize, our findings indicate that addition of data on laboratory tests and vital signs is likely to lead to improved face validity and performance of risk adjustment models. Given that these data are already part of routinely-collected patient data, the VA should consider inclusion of these data in its ongoing hospital quality measurement programs. The lack of concordance across quality measures also points to the need to identify processes of care that are more tightly linked to patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: This research has been funded by a VA HSR&D grant (IIR 08-351, A. Hanchate, PI).

References

1. Carrier E, Cross DA. Hospital Quality Reporting: Separating the Signal from the Noise. 2013
2. Boozary AS, Manchin J III, Wicker RF. The medicare hospital readmissions reduction program: Time for reform. Jama. 2015; 314:347–348. [PubMed: 26219049]
3. Esposito ML, Selker HP, Salem DN. Quantity Over Quality: How the Rise in Quality Measures is Not Producing Quality Results. J Gen Intern Med. 2015; 30:1204–7. [PubMed: 25801695]
4. Centers for Medicare & Medicaid Services. Hospital Value-Based Purchasing. 2014
5. Chatterjee P, Joynt KE. Do Cardiology Quality Measures Actually Improve Patient Outcomes? Journal of the American Heart Association. 2014; 3
6. Press MJ, Scanlon DP, Ryan AM, Zhu J, Navathe AS, Mittler JN, Volpp KG. Limits Of Readmission Rates In Measuring Hospital Quality Suggest The Need For Added Metrics. Health Affairs. 2013; 32:1083–1091. [PubMed: 23733983]
7. Abelson R. Hospital Rating Systems Differ on Best and Worst. The New York Times. 2015
8. Krumholz HM, Lin Z, Keenan PS, Chen J, Ross JS, Drye EE, Bernheim SM, Wang Y, Bradley EH, Han LF, Normand SL. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. Jama. 2013; 309:587–93. [PubMed: 23403683]
9. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, Wang Y, Normand SL, Spertus JA, Krumholz HM. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. JAMA. 2006; 296:72–8. [PubMed: 16820549]
10. Werner RM, Bradlow ET. Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. JAMA. 2006; 296:2694–2702. [PubMed: 17164455]

11. Shih T, Dimick JB. Reliability of readmission rates as a hospital quality measure in cardiac surgery. Ann Thorac Surg. 2014; 97:1214–8. [PubMed: 24492060]
12. Austin JM, Jha AK, Romano PS, Singer SJ, Vogus TJ, Wachter RM, Pronovost PJ. National hospital ratings systems share few common scores and may generate confusion instead of clarity. Health Aff (Millwood). 2015; 34:423–30. [PubMed: 25732492]
13. Iezzoni LI. The risks of risk adjustment. JAMA. 1997; 278:1600–7. [PubMed: 9370507]
14. Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. Med Care. 1999; 37:83–92. [PubMed: 10413396]
15. Lim E, Cheng Y, Reuschel C, Mbowe O, Ahn HJ, Juarez DT, Miyamura J, Seto TB, Chen JJ. Risk-Adjusted In-Hospital Mortality Models for Congestive Heart Failure and Acute Myocardial Infarction: Value of Clinical Laboratory Data and Race/Ethnicity. Health Serv Res. 2015; 50(Suppl 1):1351–71. [PubMed: 26073945]
16. Pine M, Kowlessar NM, Salemi JL, Miyamura J, Zingmond DS, Katz NE, Schindler J. Enhancing Clinical Content and Race/Ethnicity Data in Statewide Hospital Administrative Databases: Obstacles Encountered, Strategies Adopted, and Lessons Learned. Health Services Research. 2015; 50:1300–1321. [PubMed: 26119470]
17. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-Adjusting Hospital Inpatient Mortality Using Automated Inpatient, Outpatient, and Laboratory Databases. Med Care. 2008; 46:232–239. [PubMed: 18388836]
18. Pine M, Jordan HS, Elixhauser A, Fry DE, Hoaglin DC, Jones B, Meimban R, Warner D, Gonzales J. Enhancement of Claims Data to Improve Risk Adjustment of Hospital Mortality. JAMA. 2007; 297:71–76. [PubMed: 17200477]
19. Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. Med Care. 2007; 45:789–805. [PubMed: 17667314]
20. Veterans Health Administration. Hospital Report Cards. 2014
21. Bernheim SM, Wang Y, Grady JN, Bhat KR, Wang H, Abedin Z, Desai MM, Li S, Vellanky S, Lin Z, Drye E, Krumholz HM. 2011 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Mortality Measures. Submitted By Yale University/Yale-New Haven Hospital-Center for Outcomes Research & Evaluation (Yale-CORE) Prepared for Centers for Medicare & Medicaid Services (CMS). 2011
22. Render ML, Almenoff PL, Christianson A, Sales AE, Czarnecki T, Deddens JA, Freyberg RW, Eyman J, Hofer TP. A hybrid Centers for Medicaid and Medicare service mortality model in 3 diagnoses. Medical Care. 2012; 50:520–6. [PubMed: 22584887]
23. VA Information Resource Center. VIReC Research User Guide: VA Corporate Data Warehouse. 2012
24. Bernheim SM, Lin Z, Grady J, Bhat KR, Wang H, Wang Y, Abedin Z, Desai MM, Li S-X, Vellanky S, Drye EE, Krumholz HM. 2011 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures. 2011
25. DxCG Inc. DxCG RiskSmart: Clinical Classifications Guide. Boston, MA: DxCG; 2011.
26. Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation. 2010 Condition Category – ICD-9-CM Crosswalks: Acute Myocardial Infarction. 2013
27. Render ML, Almenoff PL, Christianson A, Sales AE, Czarnecki T, Deddens JA, Freyberg RW, Eyman J, Hofer TP. A hybrid Centers for Medicaid and Medicare service mortality model in 3 diagnoses. Med Care. 2012; 50:520–6. [PubMed: 22584887]
28. Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, Reed WG, Swanson TS, Ma Y, Halm EA. An Automated Model to Identify Heart Failure Patients at Risk for 30-Day Readmission or Death Using Electronic Medical Record Data. Medical Care. 2010; 48:981–988. [PubMed: 20940649]
29. Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program. 2015
30. Fonarow GCPW, Saver JL, et al. Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. JAMA: The Journal of the American Medical Association. 2012; 308:257–264. [PubMed: 22797643]

31. The Joint Commission. Facts about ORYX for Hospitals (National Hospital Quality Measures). 2013
32. Veterans Health Administration. National Surgery Office. 2013
33. Itani KMF. Fifteen years of the National Surgical Quality Improvement Program in review. The American Journal of Surgery. 2009; 198:S9–S18. [PubMed: 19874939]
34. Kelly C, Price TD. Correcting for regression to the mean in behavior and ecology. Am Nat. 2005; 166:700–7. [PubMed: 16475086]
35. Bland JM, Altman DG. Regression towards the mean. British Medical Journal. 1994; 308:1499. [PubMed: 8019287]
36. Bland JM, Altman DG. Some examples of regression towards the mean. British Medical Journal. 1994; 309:780. [PubMed: 7950567]
37. Austin PC, Reeves MJ. The relationship between the C-statistic of a risk-adjustment model and the accuracy of hospital report cards: a Monte Carlo Study. Med Care. 2013; 51:275–84. [PubMed: 23295579]
38. Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, Wang CY, Garutti RI, Jacka MJ, Sigamani A, Srinathan S, Biccard BM, Chow CK, Abraham V, Tiboni M, Pettit S, Szczeklik W, Lurati Buse G, Botto F, Guyatt G, Heels-Ansdell D, Sessler DI, Thorlund K, Garg AX, Mrkobrada M, Thomas S, Rodseth RN, Pearse RM, Thabane L, McQueen MJ, VanHelder T, Bhandari M, Bosch J, Kurz A, Polanczyk C, Malaga G, Nagele P, Le Manach Y, Leuwer M, Yusuf S. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. Jama. 2012; 307:2295–304. [PubMed: 22706835]
39. Goyal A, Spertus JA, Gosch K, Venkitachalam L, Jones PG, Van den Berghe G, Kosiborod M. Serum potassium levels and mortality in acute myocardial infarction. JAMA : the journal of the American Medical Association. 2012; 307:157–64. [PubMed: 22235086]
40. Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, Greenland P. Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA : the journal of the American Medical Association. 2010; 303:1610–6. [PubMed: 20424251]
41. Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shoshan O, Lorch SA, Volpp KG. The Hospital Compare mortality model and the volume-outcome relationship. Health Serv Res. 2010; 45:1148–67. [PubMed: 20579125]
42. Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006; 113:1683–92. [PubMed: 16549637]
43. Navathe AS, Volpp KG, Konetzka RT, Press MJ, Zhu J, Chen W, Lindrooth RC. A longitudinal analysis of the impact of hospital service line profitability on the likelihood of readmission. Med Care Res Rev. 2012; 69:414–31. [PubMed: 22466577]
44. Donabedian A. Evaluating the Quality of Medical Care. Milbank Quarterly. 1966; 44:166–203.
45. Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand SL. Variability in the measurement of hospital-wide mortality rates. The New England Journal of Medicine. 2010; 363:2530–9. [PubMed: 21175315]
46. Joynt KE, Jha AK. Thirty-Day Readmissions — Truth and Consequences. New England Journal of Medicine. 2012; 366:1366–9. [PubMed: 22455752]
47. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. 2013; 309:342–3. [PubMed: 23340629]
48. Ryan AM, Burgess JF, Pesko MF, Borden WB, Dimick JB. The Early Effects of Medicare's Mandatory Hospital Pay-for-Performance Program. Health Services Research. 2014:n/a–n/a.
49. Ryan AM, Burgess JF Jr, Tompkins CP, Wallack SS. The relationship between Medicare's process of care quality measures and mortality. Inquiry. 2009; 46:274–90. [PubMed: 19938724]
50. Byrne MM, Kuebeler M, Pietz K, Petersen LA. Effect of using information from only one system for dually eligible health care users. Med Care. 2006; 44:768–73. [PubMed: 16862039]

Table 1
Hospital-level Counts and Outcomes by Admission and Outcome Cohort

VA Hospitals, FY2007-2010

	Acute Myocardial Infarction	Heart Failure	Pneumonia
30-day Mortality Cohort			
# hospitals (n)	91	128	131
# discharges (n)	22,608	59,595	62,996
Median # discharges per hospital [range] (n)	193 [53 – 1,202]	435 [55 – 1,758]	428 [50 – 1,586]
# deaths within 30-days of admission (n)	2,202	4,695	6,431
30-day mortality rate (%)	9.7	7.9	10.2
Median 30-day mortality rate (%) per hospital [range]	9.9 [3.0 – 26.2]	8.0 [3.0 – 15.3]	9.7 [3.9 – 19.2]
30-day Readmission Cohort			
# hospitals (n)	97	130	131
# discharges (n)	25,748	78,874	69,451
Median discharges per hospital (range) (n)	233 [50 – 1,132]	575 [58 – 2,730]	471 [53 – 1,462]
# readmissions within 30-days of admission (n)	5,172	17,560	11,410
30-day readmission rate (%)	20.1	22.3	16.4
Median 30-day readmissions rate (%) per hospital (range)	20.0 [8.3 – 31.1]	22.0 [10.1 – 30.9]	15.8 [7.1 – 22.7]

Table 2

Impact of enhanced risk adjustment on model performance, by Admission and Outcome Cohort

	Acute Myocardial Infarction		Heart Failure		Pneumonia	
	Without enhancement	**With enhancement**	**Without enhancement**	**With enhancement**	**Without enhancement**	**With enhancement**
30-day Mortality Cohort						
C-statistic	0.79 [0.78-0.81]	0.85 [0.85-0.86]	0.73 [0.72-0.73]	0.81 [0.80-0.81]	0.76 [0.75-0.77]	0.82 [0.82-0.83]
Pseudo-R2	0.08 [0.07-0.08]	0.13 [0.12-0.14]	0.04 [0.03-0.04]	0.08 [0.07-0.08]	0.06 [0.06-0.07]	0.1 [0.10-0.11]
Predicted 30-day mortality rate by decile of predicted risk						
Lowest decile	0.75 [0.34-1.24]	0.58 [0.18-0.71]	1.28 [1.05-1.61]	0.65 [0.46-0.86]	1.24 [0.97-1.61]	0.56 [0.36-0.75]
Highest decile	33.32 [31.52-35.94]	43.50 [40.87-46.39]	21.95 [20.61-23.36]	31.03 [29.76-33.03]	30.45 [28.80-32.22]	39.72 [38.10-41.80]
30-day Readmission Cohort						
C-statistic	0.62 [0.61,0.63]	0.64 [0.63,0.65]	0.60 [0.60,0.61]	0.63 [0.62,0.63]	0.63 [0.63,0.64]	0.64 [0.64,0.65]
Pseudo-R2	0.03 [0.02,0.04]	0.04 [0.03,0.04]	0.02 [0.02,0.03]	0.03 [0.03,0.04]	0.03 [0.02,0.03]	0.03 [0.03,0.03]
Predicted 30-day readmission rate by decile of predicted risk						
Lowest decile	9.91 [8.39,11.51]	9.32 [8.30,10.83]	13.29 [12.20,14.34]	10.85 [9.93,11.65]	7.49 [6.68,8.28]	6.65 [5.95,7.33]
Highest decile	35.16 [32.62,38.90]	36.95 [34.12,40.30]	36.33 [34.78,37.85]	37.07 [35.94,38.95]	29.38 [28.10,30.60]	30.24 [29.08,32.07]

Note:

1) 95% confidence interval was calculated using bootstrap samples (N=500).

Table 3

Number of Hospitals Grouped by Quintiles of Risk Standardized Mortality Rate (RSMR) from Administrative and Enhanced Data Models

Without enhancement	Enhanced Model RSMR Quintiles				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
30-day Mortality Cohort					
Acute Myocardial Infarction					
Quintile 1 (lowest mortality)	12	5	1	1	0
Quintile 2	4	6	5	2	1
Quintile 3	2	4	8	4	0
Quintile 4	1	2	3	7	5
Quintile 5 (highest mortality)	0	1	1	4	12
Heart Failure					
Quintile 1 (lowest mortality)	17	8	1	0	0
Quintile 2	8	10	6	2	0
Quintile 3	1	7	11	5	1
Quintile 4	0	1	4	14	7
Quintile 5 (highest mortality)	0	0	3	5	17
Pneumonia					
Quintile 1 (lowest mortality)	21	4	0	2	0
Quintile 2	6	11	8	1	0
Quintile 3	0	6	11	7	2
Quintile 4	0	5	4	8	9
Quintile 5 (highest mortality)	0	0	3	8	15
30-day Readmission Cohort					
Acute Myocardial Infarction					
Quintile 1 (lowest readmission)	18	2	0	0	0
Quintile 2	2	15	2	0	0
Quintile 3	0	2	15	3	0
Quintile 4	0	0	3	13	3
Quintile 5 (highest readmission)	0	0	0	3	16

Without enhancement	Enhanced Model RSMR Quintiles				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Heart Failure					
Quintile 1 (lowest readmission)	22	4	0	0	0
Quintile 2	4	18	4	0	0
Quintile 3	0	4	14	8	0
Quintile 4	0	0	8	12	6
Quintile 5 (highest readmission)	0	0	0	6	20
Pneumonia					
Quintile 1 (lowest readmission)	23	4	0	0	0
Quintile 2	4	17	5	0	0
Quintile 3	0	5	16	5	0
Quintile 4	0	0	5	16	5
Quintile 5 (highest readmission)	0	0	0	5	21

Table 4

Enhanced Risk Adjustment and Concordance of 30-day Mortality and 30-day Readmission Performance

Kappa statistic [95% Confidence Interval] reported for concordance between RSMR and RSRR quartiles

	Cohort		
	AMI (n=91)	**HF (n=128)**	**Pneumonia (n=131)**
Concordance between RSMR & RSRR quartiles			
Base Model	0.05 [−0.07, 0.17]	0.09 [−0.01, 0.19]	0.17 [0.07, 0.27]
Enhanced Model	0.09 [−0.03, 0.21]	0.04 [−0.04, 0.14]	0.06 [−0.06, 0.16]
Rank correlation, Spearman correlation [p-value of test of independence of performance measures]			
Base Model	0.20 [p=0.06]	0.15 [p=0.09]	0.27 [p<0.01]
Enhanced Model	0.17[p=0.11]	0.25 [p<0.01]	0.32 [p<0.01]

Table 5

Concordance of RSMR and RSRR with Other Hospital Quality Measures: Kappa Statistic

Kappa statistic [95% Confidence Interval] reported for concordance between RSMR/RSRR quartiles and each of the Other Quality Measures

	Cohort Pair Compared		
	AMI (n=91)	**HF (n=128)**	**Pneumonia (n=131)**
30-day Mortality Cohort			
1. ORYX Process Indicators (quartiles)			
Administrative Data Model	−0.04 [−0.16, 0.08]	0.10 [0.0, 0.20]	0.02 [−0.08, 0.12]
Enhanced Data Model	−0.05 [−0.17, 0.07]	0.09 [−0.01, 0.11]	0.06 [−0.04, 0.16]
2. VASQIP Surgical Mortality Rates (quartiles)			
Administrative Data Model	0.03 [−0.09, 0.15]	0.04 [−0.08, 0.16]	0.04 [−0.08, 0.16]
Enhanced Data Model	0.02 [−0.10, 0.14]	0.11 [−0.01, 0.13]	0.08 [−0.04, 0.20]
30-day Readmission Cohort			
1. ORYX Process Indicators (quartiles)			
Administrative Data Model	−0.03 [−0.09, 0.15]	−0.03 [−0.13, 0.07]	−0.05 [−0.15, 0.05]
Enhanced Data Model	−0.01 [−0.12, 0.12]	−0.11 [−0.21, −0.01]	−0.02 [−0.14, 0.10]
2. VASQIP Surgical Mortality Rates (quartiles)			
Administrative Data Model	−0.01 [−0.13, 0.11]	0.03 [−0.09, 0.15]	0.03 [−0.09, 0.15]
Enhanced Data Model	0.04 [−0.08, 0.16]	0.13 [0.01, 0.25]	0.10 [−0.02, 0.22]