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

Amresh D. Hanchate[a,b,*], Kelly L. Stolzmann[a], Amy K. Rosen[a,c], Aaron S. Fink[d], Michael Shwartz[a,e], Arlene S. Ash[f], Hassen Abdulkerim[a], Mary Jo V. Pugh[g,h], Priti Shokeen[a], Ann Borzecki[b,i,j]

[a] Center for Healthcare Organization and Implementation Research (CHOIR), VA Boston Healthcare System, Boston, MA 02130, USA
[b] Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02118, USA
[c] Department of Surgery, Boston University School of Medicine, Boston, MA 02118, USA
[d] Professor Emeritus of Surgery, Emory University School of Medicine, Atlanta, GA 30322, USA
[e] Department of Operations and Technology Management, Boston University School of Management, Boston, MA 02215, USA
[f] Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA 01655, USA
[g] South Texas Veterans Health Care System, San Antonio, TX 78229, USA
[h] Department of Epidemiology and Biostatistics, University of Texas Health Science Center, San Antonio, TX 78229, USA
[i] Center for Healthcare Organization and Implementation Research (CHOIR), Bedford VAMC, Bedford, MA 01730, USA
[j] Department of Health Law, Policy and Management, Boston University School of Public Health, Boston, MA 02118, USA

ARTICLE INFO

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-d mortality and 30-d readmission rates as a result of 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-d mortality, data enhancement improved model performance, and significantly changed hospital performance profiles; for 30-d 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.

1. Introduction

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-d mortality with 30-d 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-d mortality and 30-d 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.

2. 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.

2.1. 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.

2.2. 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-d mortality and 30-d 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 h, 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 h after admission time, (b) extending the time window beyond 24 h 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/HCO_3) 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 h 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.

2.3. 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=1000).[21] Hospital