

The Veterans Health Affairs experience in measuring and reporting inpatient mortality

Marta L. Render, M.D., Peter Almenoff, M.D.

Introduction

The Veterans Health Affairs (VA) began a national program to measure and report risk adjusted mortality and length of stay in its intensive care units (ICU) in October of 2004 expanding to all patients admitted for acute medical or surgical conditions in October of 2008. The program, the VA Inpatient Evaluation Center (IPEC), uses a validated risk measure [1-3]. The VA IPEC uses electronic data to measure, people resources to drive change. This brief paper provides an overview of the VA risk methods and lessons learned since implementation.

Risk adjustment methodology.

The hallmark of the VA risk model is its reliance solely on electronic elements from the VA medical record. Data is directly extracted from the patient treatment file, patient file, laboratory file, and radiology file of each site using customized programming and sent encrypted to the central data repository located at the Cincinnati VA Medical Center. Diagnosis, comorbid disease

Preferred attributes	VA Risk models	
	ICU	Acute Care
Clear definition of patient sample	All patients admitted to any ICU in the VHA, defined by treating specialty in EMR	All patients admitted to acute care excluding rehab, psych, NH,
Clinical coherence of variables	Includes variables available in the EMR and included in other ICU risk models, substitutes laboratory abnormalities reflecting variation in organ perfusion for physiology variables	
Sufficiently high quality and timely data	Data directed extracted from hospital computer system, reported quarterly	
Appropriate reference time before which covariates derived and after which outcome occurs	Lab values from 24 hours surrounding ICU admission; Outcome hospital and 30-day mortality	Lab values from 24 hours surrounding admission to hospital, outcome - 30 day mortality after admission to hospital
Application of analytic approach accounting for nested data	SMR's and 95% confidence intervals created using 2 level random effects model	SMR's and 95% confidence intervals created using 2 level random effects model
Methodology disclosure	Published models and updates	

burden, source of admission (emergency room/ outpatient clinic, operating room, nursing home, ward, or outside hospital) are included as categorical variables, while age and the worst measured value of 11 laboratory tests (sodium, blood urea nitrogen, creatinine, bilirubin, albumin, glucose, hematocrit, white blood cell count, pH/ paCO₂, and PaO₂) are treated as cubic splines. Patients are assigned to one mutually exclusive diagnosis using ICD-9-CM coding from the ICU bedsection. Comorbid disease burden is assessed using Elixhauser's approach [4, 5]. The risk model in intensive care unit patients has excellent calibration and discrimination when used to predict hospital mortality (validation set of 220,813 cases: c statistic 0.892, Hosmer Lemeshow Chi square 376) or mortality at 30 days (validation set of 193,944: C statistic 0.87, Hosmer Lemeshow goodness of fit statistic chi square 161). Advantages of use of a wholly electronic dataset includes access to reliable laboratory values, improved face validity, reduced cost compared to manual data extraction with the attendant opportunity to extend mortality measurement across an entire system, and ease of updating the models.

Measures derived from mortality data

From the risk model, for each intensive care or hospital, a standardized mortality ratio (SMR, observed / predicted deaths) is determined using a 2-level hierarchical random effects model for outcomes of death at hospital discharge and death at 30 days from admission. The random effects model accounts for nesting of data in each ICU or hospital to improve estimate accuracy. An observed minus predicted length of stay is also determined. Use of the difference between predicted and observed length of stay allows

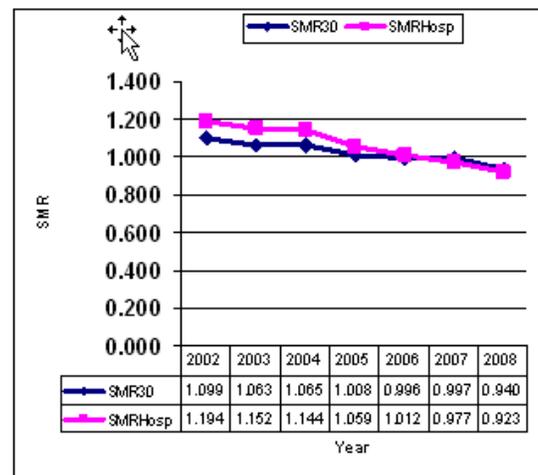
estimation of the cost avoidance or opportunity loss when a unit OMELOS is multiplied by the daily cost of an ICU day and the annual census. We also now track unadjusted mortality at hospital discharge and 30 days for both the acute care and ICU patients and have piloted the unadjusted mortality of patients transferred from the ward to the ICU as an indicator of the ability of a hospital to detect and rescue deteriorating patients. Finally, using the same risk model, we created a physiologic case mix index where the numerator was the predicted mortality for patients in the specific ICU and the denominator the predicted mortality for patients in all VA ICUs. Because the proportion of operative cases (those with surgery in the 24 hours surrounding ICU admission) significantly influenced results when aggregated with non-operative cases, case mix indices were created separately for operative and non-operative and then a weighted measure based on the proportion of each was determined.

Results

The coefficients or weights of the predictors of the model, developed on a pilot dataset from 2002 – 2004, were fixed to allow tracking of ICU performance overtime, and SMR drifted significantly downward. This drift made interpretation of the SMR more difficult. For example in 2007, the VA SMR nationally was 0.8.

Hospitals or ICUs with SMR only slightly above 1 then appeared on face to be “average” (where SMR of 1 = observed/ predicted deaths) but in fact might have a 20% difference in risk adjusted mortality. To avoid confusion, the risk models are now recalibrated at the

Figure 1. VA SMR30 and SMR hosp



beginning of each year on the prior two years data, and the fixed weights of the predictors then applied to the new year as well as the prior years (to 2002). The reason for the downward drift is unknown, although temporally related to VA initiatives to improve implementation of evidenced based practices. Standardized mortality ratios varied somewhat based on the type of the ICU and level of complexity of the ICU. The ability to stratify by type or level of ICU and create benchmarks that were ICU specific improved the early face validity and acceptance of this measurement approach.

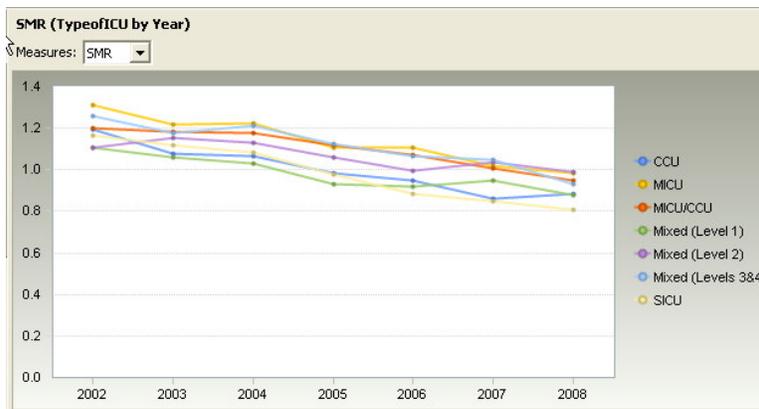


Figure 2. Variation in SMR stratified by type of ICU and level of complexity

In some ICUs, there was a dramatic difference between the SMR that predicted death at hospital discharge (SMR_{hosp}) from the SMR predicting death at 30 days (SMR_{30}).

Anecdotal follow-up suggested that variation in discharge practices related to the availability of long term acute care units and palliative care units were important in hospitals with large differences when their SMR_{hosp} was subtracted from SMR_{30} .

Variation in unadjusted mortality of patients transferred from the ward to the ICU also varied significantly (2004 : overall 20%, range 6- 36%, 2008: overall 16%, range 4- 32%) and has fallen across the VA coincident with implementation of rapid response teams.

Use of multiple mortality measures improves confidence in using the results. For instance, a small ICU with a higher SMR at hospital discharge and a normal SMR at 30 days likely has “normal” performance” and the high SMR_{hosp} is

	Unadj		Unadj 30d	Unadj TX		
	SMR	SMR30	Hosp Mort	Mort	Mort	Case Mix
	2006	2006	2006	2006	2006	2006
Hosp 1	0.989	1.041	10.54%	12.49%	25.16%	1.318
Hosp 2	0.822	0.681	8.85%	7.87%	16.67%	1.401
Hosp 3	1.055	1.038	13.04%	13.85%	29.67%	1.456
Hosp 4	1.158	0.997	11.43%	11.18%	33.33%	0.900
Hosp 5	1.332	1.291	12.42%	13.80%	24.18%	1.056
	2007	2007	2007	2007	2007	2007
Hosp 1	0.686	0.680	7.50%	8.25%	15.41%	1.469
Hosp 2	0.683	0.701	7.04%	7.95%	16.23%	1.396
Hosp 3	0.976	0.932	10.67%	10.75%	26.14%	1.430
Hosp 4	1.244	1.031	11.52%	10.53%	30.36%	0.970
Hosp 5	1.299	1.257	12.62%	13.93%	23.60%	1.101
	2008	2008	2008	2008	2008	2008
Hosp 1	0.751	0.735	8.73%	9.57%	16.89%	1.365
Hosp 2	0.509	0.497	5.26%	5.76%	12.09%	1.303
Hosp 3	0.691	0.784	7.40%	9.39%	17.17%	1.296
Hosp 4	0.824	0.770	8.20%	9.62%	16.90%	0.953
Hosp 5	1.199	1.172	11.95%	14.02%	20.21%	1.173

related to limited resource for long term acute care patients (on vents, severe debilitation). When death occurs in a patient at this hospital even after a year of inpatient care, it counted toward that hospital’s mortality; while similar patients in other hospitals were “discharged” when sent to long term acute care hospitals or rehab facilities, and counted as survivors. The signal to noise ratio appeared improved when multiple mortality measures were tracked and concordant (SMR at hospital discharge, unadjusted mortality, mortality at transfer to the ICU from the ward). Similarly, given an imperfect model, when VA case mix was low and SMR elevated again the signal increased. The VA case mix also allowed tracking of the relative severity of illness of hospitalized patients in a system with hospitals with varying services and complexity.

Lessons learned

Following 4 intense years of building a system that measures and reports risk adjusted mortality in 138 hospitals nationally, we have some lessons regarding structure of a national measurement system outside of the VA that might be valid. First, a risk adjustment model that predicts death at 30-days in addition to a model predicting death at

hospital discharge will be important to avoid gaming. Next, resources and expertise to support recalibration of the weights of the model at appropriate time frame, using a large dataset will be needed as part of the infrastructure of the program. Use of laboratory data which provides a surrogate for variation in physiology will likely 1) improve face validity, 2) is probably possible now given the use of computerized systems for laboratory data retrieval in most hospitals, and 3) likely neutralizes the impact of gaming using administrative data. Fourth, regionalization of the results (as has been done with Healthcompare) and/ or stratification by mission or complexity of facility will improve the usability of the results. Finally, because all performance measures inherently will be gamed, thinking about mortality measures as a bundle of indicators rather than a single gold standard might improve the information created by the models.

References

1. Render, M.L., et al., *Automated computerized intensive care unit severity of illness measure in the Department of Veterans Affairs: preliminary results. SISVista Investigators. Scrutiny of ICU Severity Veterans Health Systems Technology Architecture*. Crit Care Med, 2000. **28**(10): p. 3540-6.
2. Render, M.L., et al., *Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure*. Crit Care Med, 2005. **33**(5): p. 930-9.
3. Render, M.L., et al., *Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration*. Crit Care Med, 2008. **36**(4): p. 1031-42.
4. Elixhauser, A., et al., *Comorbidity measures for use with administrative data*. Med Care, 1998. **36**(1): p. 8-27.
5. Johnston, J.A., et al., *Impact of different measures of comorbid disease on predicted mortality of intensive care unit patients*. Med Care, 2002. **40**(10): p. 929-40.