Published in final edited form as:

Crit Care Med. 2017 November ; 45(11): 1863–1870. doi:10.1097/CCM.0000000000002651.

Adaptation of the Acute Organ Failure Score for Use in a Medicare population

Katherine R. Courtright, MD, MS^{1,2}, Scott D. Halpern, MD, PhD^{1,2,3}, Brian Bayes, MS³, Michael O. Harhay, PhD, MPH³, Eli Raneses, MPH⁴, Patricia Kipnis, PhD⁵, Gabriel J. Escobar, MD⁵, and Meeta Prasad Kerlin, MD, MSCE^{1,2,3}

¹Division of Pulmonary, Allergy & Critical Care Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia ²Leonard Davis Institute of Health Economics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA ³Center for Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia ⁴F. Edward Hébert School of Medicine, Uniform Services University of the Health Sciences, Bethesda Maryland ⁵Division of Research, Kaiser Permanente Northern California, Oakland, CA

Abstract

Objective—Without widely available physiologic data, a need exists for ICU risk adjustment methods that can be applied to administrative data. We sought to expand the generalizability of the Acute Organ Failure Score (AOFS) by adapting it to a commonly used administrative database..

Design—Retrospective cohort study

Setting—151 hospitals in Pennsylvania

Patients—90,733 ICU admissions among 77,040 unique patients between January 1, 2009 and December 1, 2009 in the Medicare Provider Analysis and Review database.

Measurements and Main Results—We used multivariable logistic regression on a random split cohort to predict 30-day mortality, and to examine the impact of using different comorbidity measures in the model and adding historical claims data. Overall 30-day mortality was 17.6%. In the validation cohort, using the original AOFS model's β coefficients resulted in poor discrimination (C-statistic 0.644; 95% confidence interval [CI]: 0.639–0.649). The model's C-statistic improved to 0.721 (95% CI: 0.711–0.730) when the Medicare cohort was used to re-calibrate the β coefficients. Model discrimination improved further when comorbidity was expressed as the COMorbidity Point Score 2 (C-statistic 0.737; 95% CI: 0.728–0.747; $p < 0.001$) or the Elixhauser index (C-statistic 0.748; 95% CI: 0.739–0.757) instead of the Charlson index. Adding historical claims data increased the number of comorbidities identified, but did not enhance model performance.

Conclusions—Modification of the AOFS resulted in good model discrimination among a diverse population regardless of comorbidity measure used. This study expands the use of the

Corresponding author: Katherine Courtright, MD MS, 5048 Gates Building, 3400 Spruce Street, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, katherine.courtright@uphs.upenn.edu, Phone: 215-662-7906.

Conflicts of Interest: All authors have disclosed that they do not have any conflicts of interest.

AOFS for risk adjustment in ICU research and outcomes reporting using standard administrative data.

Keywords

intensive care; risk adjustment; comorbidity; international classification of diseases; administrative; mortality prediction

Introduction

Fair comparisons of mortality rates across hospitals and among clinicians rely on accurate methods to adjust for variation in patients' severities of illness. Similar adjustment is the bedrock of observational research designed to improve upon the quality of care. These considerations are particularly important when evaluating patients admitted to ICUs given their high rates of death and multiple competing risks for mortality. Although there are numerous well-validated ICU risk adjustment methodologies [1–6], they require physiological data that are often lacking in administrative datasets, which offer the large sample sizes needed to show statistically and clinically meaningful differences in mortality [7, 8] and are commonly used for hospital performance measurement and reporting [9, 10].

Elias et al recently developed the Acute Organ Failure Score (AOFS) for predicting intensive care unit (ICU) mortality using exclusively administrative data [11]. This methodology incorporates acute organ failures identified by a combination of *International Classification of Diseases*, 9th Revision, Clinical Modification (ICD-9-CM) codes [12], in addition to patient demographics and comorbidities according to the Charlson comorbidity index (CCI). The AOFS showed good performance characteristics in derivation and validation cohorts from separate hospitals. However, the model was developed using a data registry that contained temporal details of diagnosis codes and processes of care that are often absent from standard, widely used administrative databases, thus limiting its generalizability. Moreover, the AOFS was developed and validated in cohorts unique to a single, local health system. External validation in a distinct and broadly representative population from a common data source is essential to demonstrate true utility of a prediction model [13]. Finally, among the several commonly used claims-based comorbidity measures, the Charlson comorbidity index is the least comprehensive and may have disadvantaged the potential discrimination of the AOFS [14–16].

In this study, we sought to expand the generalizability of the AOFS for use in standard claims data via three mechanisms: modification of model covariates to accommodate standard administrative data; use of historical comorbidity claims and more granular comorbidity summary measures to enhance predictive accuracy; and inclusion of a diverse critically ill population to improve external validity.

Materials and Methods

Patients and data sources

We performed a retrospective cohort study using the Centers for Medicare and Medicaid Services (CMS) Medicare Provider Analysis and Review (MedPAR) files[17], an

administrative dataset containing hospital discharge data for Medicare beneficiaries. We linked the Medicare Beneficiary Summary and Carrier files to ascertain date of death and outpatient provider-based claims, respectively.

We included all beneficiaries ≥ 18 years who had an ICU stay ≥ 1 day (defined by the presence of an ICU revenue center code, excluding intermediate or psychiatry ICUs) [18] in an acute care hospital in Pennsylvania between January 1, 2009 and December 1, 2009. We considered multiple hospitalizations for the same patient to be a single episode if the second admission date occurred within one calendar day of the first hospital discharge date [19].

The Pennsylvania Health Care Cost Containment Council (PHC4) is a nonprofit government agency that reports outcomes data on all patients discharged from nonfederal hospitals in Pennsylvania. 2009 was the final year in which PHC4 used a physiology-based risk adjustment method, thus we linked MedPAR with PHC4 inpatient discharge data to obtain the predicted probability of in-hospital death according to the MediQual score [9].

The Acute Organ Failure Score ICU risk adjustment methodology

Details of the AOFS risk adjustment methodology are provided elsewhere [11]. Briefly, the AOFS was developed and validated using a registry of 92,886 patients who received critical care at two urban hospitals. Eligible patients were identified by the presence of current procedural terminology (CPT) code 99291 for provision of critical care services. The final model included age, diagnostic-related group (DRG) classification [20], race, the CCI [21], an indicator for sepsis [12], and 6 types of acute organ failure (respiratory, renal, hepatic, hematologic, metabolic, or neurologic) [11, 12]. ICD-9-CM codes indicating acute organ failure were included if assigned within the two days before or after the earliest CPT code 99291. Multivariate logistic regression was performed and an integer risk score was calculated based on a linear transformation of the corresponding β regression coefficient. The final risk score had a C-statistic of 0.736.

Modification of Variables and Definitions

Variables were defined according to the AOFS methodology [11] with several exceptions. ICU admission was determined by the MedPAR “ICU indicator code” variable that is based on revenue center codes and denotes admission to an ICU instead of the less specific CPT code, which indicates the provision of critical care services regardless of service location. Because the timing of ICU admission during the hospitalization is unavailable in MedPAR data, we had to modify the time qualifier for acute diagnoses such that we included sepsis and the six acute organ failures only if the associated ICD-9-CM codes were designated as present on hospital admission (POA). This conservative approach avoided counting diagnoses that may have occurred during or after the ICU stay. Likewise, the primary outcome for this study was all-cause 30-day hospital mortality, that is, death within 30 days of hospital admission rather than ICU admission.

For all comorbidity measures, diagnoses were included only if designated as POA for a given hospital episode. In addition to the CCI used in the original AOFS, we also calculated a Comorbidity Point Score 2 (COPS2) and an Elixhauser index [22]. COPS2 is a weighted continuous comorbidity score developed at Kaiser Permanente for inpatient mortality risk

adjustment [23]. We used the diagnostic coding schema by Quan et al to determine the Elixhauser index [14].

We compared the performance of the modified AOFS to the MediQual score, a physiology-based mortality risk adjustment model used commonly for hospital performance comparisons [2, 9]. The MediQual probability of in-hospital death is calculated by a proprietary algorithm using data available in the initial 24 hours of hospitalization, including diagnosis and procedure codes, age, gender, and clinical lab data [2, 24, 25]. MediQual scores were calculated on patients discharged from a Pennsylvania hospital in 2009 with at least 1 of 35 conditions [26] and were unavailable if patient data were missing.

Analysis

We generated four models to predict 30-day mortality using split-validation with a random 80% of admissions for derivation and the remaining 20% for validation. We first replicated the original AOFS model using β coefficients derived from the original derivation cohort [11]. Based on a calibration-in-the-large <0 indicating systematic miscalibration and a calibration slope <1 indicating overfitting of the model in the validation cohort, we then re-estimated the β coefficients (slope) and α (intercept) using logistic regression in our derivation cohort [27]. For the other two models, we used the modified AOFS model and substituted the CCI with either the COPS2 score or the Elixhauser index.

The logit for the original AOFS model was calculated by summing the model intercept (-2.76107 ; K.B. Christopher, MD, written communication, May 2015) and β coefficients [11] multiplied by this study cohort's covariates. The expected probability of death within 30 days of hospital admission was calculated as the reciprocal of $1 + e^{-(\text{logit})}$. For all other models, the primary outcome of 30-day hospital mortality was analyzed using multivariable logistic regression. We used the predicted probabilities for all model assessments instead of transforming them into a weighted integer risk score as described by Elias et al [11]. The relative weight of each covariate in the model may vary depending on the population under study, thus limiting generalizability of such a score. Moreover, a point-based transformation is not feasible with a continuous variable (COPS2) or 31 binary variables (Elixhauser).

The unit of analysis was the hospital episode for all primary analyses. We assumed that repeated hospitalizations by an individual patient were independent because point estimates are known to be insensitive to misspecification of the random effects distribution [28] and our intent was to obtain the point estimates of probability of mortality for a given hospitalization. However, to explore the effect of multiple admissions, we conducted a sensitivity analysis restricted to index admissions. In additional sensitivity analyses, we included a 1-year look back (days 1–365 preceding hospitalization) of inpatient POA and outpatient provider-based diagnostic claims for each of the three comorbidity measures. Including additional comorbidity data sources has been shown to improve explanatory power in hospital mortality prediction models [29, 30]. Because some patients, such as Medicare Advantage members, do not have provider claims filed with Medicare, look back analyses excluded admissions without associated provider-based inpatient claims in the Carrier file, assuming that their outpatient data would be similarly incomplete.

We measured model discrimination using the area under the receiver operating characteristic curve (C-statistic) with 95% confidence intervals (CIs) for each model [31]. We assessed model calibration with the Hosmer-Lemeshow (H-L) chi-square goodness of fit test statistic [32] and visually with a calibration plot. Given the H-L test is sensitive to larger sample sizes, we also assessed random subsamples ($n=1,000$) from each cohort [33]. We additionally plotted observed vs. expected mortality rates after the latter was used to group patients in 10% risk intervals. Within each risk stratum, the expected mortality rate was deemed significantly different from the observed rate if it was outside of the 95% CI of the observed mortality rate calculated using exact methods [34]. Finally, in a subset of patients not included in the derivation cohort and for whom a MediQual score was available, we compared the areas under the curve between the MediQual score and each model using a chi-square test.

We considered two-tailed p values below 0.05 statistically significant. All analyses were performed using STATA v14.0 (StataCorp, College Station, TX). Approval to conduct this study with a waiver of the requirement for informed consent was granted by the University of Pennsylvania Institutional Review Board (Protocol #816394).

Results

During the study period, 90,733 hospital episodes among 77,040 unique patients at 151 hospitals met our inclusion criteria (Table 1). There were no differences in patient characteristics between the derivation ($n=72,586$) and validation cohorts ($n=18,147$) (Supplemental Digital Content 1). Compared to the original study cohort, our cohort had greater proportions of older, female, and white patients with fewer documented comorbidities and acute organ failures. Our cohort also had a shorter mean length of stay (LOS) (9.9 vs. 12 days; $p<0.001$) and higher 30-day mortality (17.6% vs. 12.6%; $p<0.001$).

Table 2 presents the discrimination for each model in the validation cohort. The original AOFS model discriminated poorly between 30-day survival and death among Medicare patients (C-statistic 0.644; 95% CI: 0.639–0.649). This model was also poorly calibrated to expected mortality rates (H-L chi-square 610.2; $p<0.001$) in all but one risk decile (Supplemental Digital Content 2) and overestimated observed mortality in all but the two most extreme risk strata (Fig. 1A).

The β coefficients for each modified AOFS model can be found in the supplemental data (Supplemental Digital Content 3, 4, and 5). The modified AOFS model showed improved discrimination (C-statistic 0.721; 95% CI: 0.711–0.730) and calibration (H-L chi-square 31.9; $p=0.004$), including good model fit in three random subsamples of $n=1,000$ (all p -values >0.05 ; Supplemental Digital Content 6). Predicted mortality rates differed significantly from observed rates in three risk deciles (Supplemental Digital Content Fig. 2) and overestimated risk in the 70–80% risk stratum (Fig 1B).

Discrimination of the modified AOFS further improved when comorbidity was expressed using either COPS2 (C-statistic 0.737, 95% CI: 0.728–0.747) or the Elixhauser index (C-statistic 0.748; 95% CI: 0.739–0.757). Good calibration was maintained with each

comorbidity measure (COPS2: H-L chi-square 22.8; $p=0.01$) and (Elixhauser: H-L chi-square 26.3; $p=0.003$), including in each of three smaller random subsamples from the derivation and validation cohorts (all H-L chi-square p -values >0.10 ; Supplemental Digital Content 6). Predicted mortality rates from the modified AOFS with either COPS2 or Elixhauser differed from observed mortality rates in only two risk deciles (Supplemental Digital Content 2) and overestimated risk in the 40–50% and 70–80% risk strata (Figs. 1C and 1D).

Including a 1-year look back of diagnoses from hospital and outpatient provider claims data increased the number of comorbidities identified, but did not improve the accuracy of the modified AOFS regardless of which comorbidity measure was used (Table 3). Likewise, restricting to patient's index admission did not alter performance of the modified AOFS (C-statistic 0.723; 95% CI: 0.712–0.733).

In the validation cohort, 15,975 (88.0%) admissions had an associated MediQual score, which had better discriminative ability than all AOFS models (0.771; 95% CI: 0.762–0.781; p -value for all pairwise comparisons was <0.001). However, calibration of the MediQual score was very poor (HL statistic 2537.1, $p<0.001$), and predicted mortality rates were significantly different than observed in 8 of 10 risk strata (Supplemental Digital Content 7).

Discussion

In this study, we adapted the claims-based AOFS mortality risk adjustment tool for use in a standard administrative dataset. The modified AOFS had good discrimination and calibration among an ICU population distinct from the original study cohort. The accuracy of the modified AOFS was modestly improved with a more granular comorbidity measure, but not by including additional historical comorbidity data. Finally, the modified AOFS predicted 30-day mortality among ICU patients more reliably across the spectra of risk than the MediQual physiology-based risk adjustment method.

Risk adjusted mortality is a common metric for hospital comparison in an era focused intensely on healthcare quality. Thus, accurate and reliable risk adjust methods are essential, as is a sample size sufficiently large enough to detect mortality differences. For what it lacks in clinical detail, administrative data offers researchers a large and diverse sample of individuals often followed over a prolonged period of time. The original AOFS is a novel claims-based risk adjustment model, but its use was limited to data that specified exact timing of processes of care and diagnoses during a hospitalization, which are frequently missing from or inaccurate in standard administrative data. This study successfully adapted the AOFS for use in Medicare's inpatient records file, one of the most widely used administrative databases for research and benchmarking purposes.

We took a deliberately conservative approach in adapting the AOFS for the variables available in standard administrative data. For example, we used hospital admission date to define the temporal context for the primary outcome, whereas the development of the original AOFS focused on the ICU admission date. The difference in 30-day ICU mortality in the cohort used for development of the original AOFS and this study cohort's 30-day

hospital mortality may have contributed to the systematic miscalibration seen in our validation cohort [35]. We therefore re-calibrated the model to predict 30-day hospital mortality outcome, one of the most common metrics to compare hospital performance [10].

We also included only acute diagnoses designated as POA in the risk prediction model to ensure they preceded ICU admission. However, this approach likely resulted in an underestimate of the incidence of sepsis and acute organ failure. Given the relatively high impact of these diagnoses on mortality risk as shown by the standardized point score from the original AOFS [11], this may have diminished the discriminative ability of the modified AOFS. Likewise, comorbidities may have been underestimated in this cohort of Medicare beneficiaries from calendar year 2009. Only 3 months prior, CMS began providing reimbursement reductions for several hospital-acquired conditions, including several infections often associated with sepsis [36], which is insufficient time to reflect improvement in documentation and coding practices. Despite these issues, the modified AOFS maintained good discrimination and calibration.

We demonstrated only modest improvements in the discrimination of the modified AOFS with use of more comprehensive comorbidity metrics. This approach was based on existing evidence that the Elixhauser index is superior at predicting hospital mortality among a general adult inpatient population [16]. The COPS2 model, a key component of the highly predictive Kaiser Permanente hospital mortality risk adjustment method [23, 37], is more comprehensive still. This study does not assess if the incremental improvement in c-statistic between the models translates to a meaningful difference for purposes of outcomes research or clinical decision-making. Importantly, CCI, COPS2 and the Elixhauser index can all be calculated using freely available resources [14, 23], thus maintaining the usability and generalizability of the modified AOFS regardless of which metric is chosen.

We attempted to further improve upon the performance of the modified AOFS by supplementing inpatient POA comorbidities with those captured from outpatient claims. When used in isolation, the prognostic ability of standard comorbidity indices appears to benefit from inclusion of a variety of data sources [16]. However, adding historical data did not improve our model performance despite a measurable increase in the capture of comorbidities among our population. This finding may be explained, in part, by the fact that when combined with acute physiological parameters, chronic conditions have relatively less influence on predictive accuracy of risk adjustment models among critically ill patients [11, 38, 39].

In the original study, the AOFS and APACHE II score performed comparably for predicting 30-day ICU mortality [11]. In this study's cohort, the MediQual score was better at discriminating among patients who survived vs. died than the modified AOFS. This finding is consistent with multiple prior studies demonstrating that physiological data improves discrimination of predictive models [23, 38, 40, 41]. However, the modified AOFS resulted in better calibration – that is, agreement between observed and expected risk predictions – than the MediQual score. The proprietary nature of the MediQual score limits our ability to explore potential mechanisms for this difference. Regardless, good model calibration is plausibly more important than discrimination for purposes of mortality predictions, which

inform shared decision-making and accurate risk adjustment in clinical practice and observational research, respectively.

This study has several important potential limitations. This cohort comprised Medicare patients in Pennsylvania only, which may limit the model's generalizability to younger populations or those in countries with different ICU admission practices and mortality rates as compared to the U.S. [42]. Although current critical care delivery and coding practices may have changed over the past decade [43], we were reassured of the robustness of the model to potential temporal trends given that the performance of our modified AOFS model was the same or better as that found in the original AOFS study that used data spanning two decades. However, future efforts should include an assessment of the modified AOFS in a more contemporary ICU cohort and using updated coding schema. Although widely used, the c-statistic is insensitive to model fit and does not have a directly clinically relevant interpretation [44], thus we included multiple calibration metrics in our analysis. Finally, our modification to use only POA acute organ failure diagnoses likely resulted in missed relevant cases of sepsis and acute organ failures, which may have reduced the model's performance in this cohort.

Conclusion

This study successfully adapted the AOFS for use in a commonly used administrative database, and modestly improved model performance by using a more granular comorbidity metric. Although inferior to predictive models containing detailed physiological data, the modified AOFS provides a claims-based risk adjustment method to facilitate the study of important population-level questions about the delivery and quality of critical care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank PHC4 for the provision of data for this research study. PHC4, its agents, and staff, have made no representation, guarantee, or warranty, express or implied, that the data are error-free, or that the use of the data will avoid differences of opinion or interpretation. The analysis was not prepared by PHC4. The analysis was done by the authors at the University of Pennsylvania. PHC4, its agents and staff, bear no responsibility or liability for the results of the analysis, which are solely the opinion of the study authors.

Source of Funding: This work was supported in part by a grant from the National Heart, Lung, and Blood Institute K08HL11677103 (MPK). The funding source played no role in the design and conduct of the study, collection, management, analysis, and interpretation of the data, or preparation, review, or approval of the manuscript.

Copyright form disclosure: Drs. Courtright, Halpern, Bayes, Harhay, and Kerlin received support for article research from the National Institutes of Health (NIH). Dr. Bayes's institution received funding from the NIH. Dr. Escobar's institution received funding from the Gordon and Betty Moore Foundation (for work on predictive modeling for in-hospital deterioration and rehospitalization) and from Merck (for research project on recurrent *Clostridium difficile* infection).

References

1. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991; 100:1619–1636. [PubMed: 1959406]
2. Iezzoni LI, Moskowitz MA. A clinical assessment of MedisGroups. *JAMA*. 1988; 260:3159–3163. [PubMed: 3184394]
3. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med*. 1996; 22:707–710. [PubMed: 8844239]
4. Metnitz PG, Moreno RP, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med*. 2005; 31:1336–1344. [PubMed: 16132893]
5. Moreno RP, Metnitz PG, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med*. 2005; 31:1345–1355. [PubMed: 16132892]
6. Higgins TL, Kramer AA, Nathanson BH, et al. Prospective validation of the intensive care unit admission Mortality Probability Model (MPM0-III). *Crit Care Med*. 2009; 37:1619–1623. [PubMed: 19325480]
7. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA*. 2004; 292:847–851. [PubMed: 15315999]
8. Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care*. 1996; 34:737–753. [PubMed: 8709656]
9. [Accessed Sept 8, 2016] Pennsylvania Health Care Cost Containment Council: Public Reports. Available at <http://www.phc4.org/reports/>
10. Centers for Medicare & Medicaid Services. [Accessed Sept 8, 2016] Medicare Hospital Compare. Available at <https://www.medicare.gov/hospitalcompare/search.html>
11. Elias KM, Moromizato T, Gibbons FK, et al. Derivation and validation of the acute organ failure score to predict outcome in critically ill patients: a cohort study. *Crit Care Med*. 2015; 43:856–864. [PubMed: 25746746]
12. Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med*. 2003; 348:1546–1554. [PubMed: 12700374]
13. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J of Clin Epidemiol*. 2015; 69:245–247. [PubMed: 25981519]
14. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005; 43:1130–1139. [PubMed: 16224307]
15. van Walraven C, Escobar GJ, Greene JD, et al. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. *J of Clin Epidemiol*. 2010; 63:798–803. [PubMed: 20004550]
16. Yurkovich M, Avina-Zubieta JA, Thomas J, et al. A systematic review identifies valid comorbidity indices derived from administrative health data. *J of Clin Epidemiol*. 2015; 68:3–14. [PubMed: 25441702]
17. Centers for Medicare & Medicaid Services. [Accessed Aug 16, 2016] Medicare Provider Analysis and Review (MEDPAR) file. Available at <http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/index.html>
18. Weissman GE, Hubbard RA, Kohn R, et al. 2017 Validation of an administrative definition of intensive care unit admission using revenue center codes. *Crit Care Med*.
19. Iwashyna TJ, Christie JD, Moody J, et al. The structure of critical care transfer networks. *Med Care*. 2009; 47:787–793. [PubMed: 19536030]
20. Palmer G, Reid B. Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. *Health Serv Manage Res*. 2001; 14:71–81. [PubMed: 11373998]
21. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J of Clin Epidemiol*. 1992; 45:613–619. [PubMed: 1607900]

22. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998; 36:8–27. [PubMed: 9431328]
23. Escobar GJ, Gardner MN, Greene JD, et al. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care*. 2013; 51:446–453. [PubMed: 23579354]
24. Iezzoni LI. The risks of risk adjustment. *JAMA*. 1997; 278:1600–1607. [PubMed: 9370507]
25. Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. *Chest*. 2008; 133:1319–1327. [PubMed: 18403657]
26. [Accessed June 13 2017] Pennsylvania Health Care Cost Containment Council: 35 Diseases, Procedures, and Medical Conditions for which Laboratory Data is Required. Available at <http://www.phc4.org/dept/dc/docs/list35.pdf>
27. Steyerberg, EW. Updating for a New Setting. In: Gail, M, Krickeberg, K, Samet, J, Tsatis, A, Wong, W, editors. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer Science; 2009. 361–389.
28. Butler SM, Louis TA. Random effects models with non-parametric priors. *Stat Med*. 1992; 11:1981–2000. [PubMed: 1480884]
29. Zhang JX, Iwashyna TJ, Christakis NA. The performance of different lookback periods and sources of information for Charlson comorbidity adjustment in Medicare claims. *Med Care*. 1999; 37:1128–1139. [PubMed: 10549615]
30. Preen DB, Holman CD, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. *J of Clin Epidemiol*. 2006; 59:940–946. [PubMed: 16895817]
31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
32. Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997; 16:965–980. [PubMed: 9160492]
33. Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Stat Med*. 2013; 32:67–80. [PubMed: 22833304]
34. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26:404–413.
35. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010; 21:128–138. [PubMed: 20010215]
36. Centers for Medicare & Medicaid Services. [Accessed May 3, 2017] Hospital-Acquired Conditions: Present on Admission Indicator. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/index.html>
37. Escobar GJ, Greene JD, Scheirer P, et al. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care*. 2008; 46:232–239. [PubMed: 18388836]
38. Needham DM, Scales DC, Laupacis A, et al. A systematic review of the Charlson comorbidity index using Canadian administrative databases: a perspective on risk adjustment in critical care research. *J Crit Care*. 2005; 20:12–19. [PubMed: 16015512]
39. Poses RM, McClish DK, Smith WR, et al. Prediction of survival of critically ill patients by admission comorbidity. *J of Clin Epidemiol*. 1996; 49:743–747. [PubMed: 8691223]
40. Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA*. 2007; 297:71–76. [PubMed: 17200477]
41. Render ML, Kim HM, Welsh DE, et al. Automated intensive care unit risk adjustment: results from a National Veterans Affairs study. *Crit Care Med*. 2003; 31:1638–1646. [PubMed: 12794398]
42. Wunsch H, Angus DC, Harrison DA, et al. Comparison of medical admissions to intensive care units in the United States and United Kingdom. *Am J Resp Crit Care*. 2011; 183:1666–1673.
43. Kohn R, Madden V, Kahn JM, et al. Diffusion of Evidence-based Intensive Care Unit Organizational Practices. A State-Wide Analysis. *Ann Am Thorac Soc*. 2017; 14:254–61. [PubMed: 28076685]

44. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115:928–935. [PubMed: 17309939]

Fig 1.

These figures represent the goodness-of-fit of each Acute Organ Failure Score (AOFS) model predicting 30-day mortality by expected risk strata. (A) Original AOFS β coefficients. (B) Modified AOFS β coefficients. (C) AOFS with comorbidity determined by COPS2. (D) AOFS with comorbidity determined by Elixhauser Index. For each model, the validation cohort was grouped by expected mortality into 10% categories. The total number of people in each category is presented by the bar graph (right vertical axis). Each category presents the expected (light blue line) and observed (dark blue line) mortalities (left vertical axis). Within each risk strata, expected mortality rates were deemed significantly different from observed rates if they were outside of the 95% confidence interval of the observed mortality rate that was calculated using exact methods [33].

Table 1

Comparison of original and present study patient cohorts

Measure ^a	Elias et al.	Present study
Study period	Nov 1997 to Feb 2011	Jan 1 to Dec 1 2009
ICU admissions, (<i>n</i>)	92,866	90,733
Patients, (<i>n</i>)	92,886	77,040
Hospitals, (<i>n</i>)	2	151
Patient age (years), mean (SD)	61.6 (17.7)	74.2 (12.2)
Male, <i>n</i> (%)	53,658 (57.7)	44,207 (48.7)
Nonwhite race <i>n</i> (%)	18,762 (20.2)	10,406 (11.5)
Patient type, <i>n</i> (%)		
Medical	46,683 (50.3)	52,789 (58.2)
Charlson Comorbidity Index, <i>n</i> (%)		
0 – 1	35,277 (38.0)	47,327 (52.2)
2 – 3	31,587 (34.0)	31,880 (35.1)
4 – 6	20,907 (22.5)	9,505 (10.5)
7	5,115 (5.5)	2,021 (2.2)
Elixhauser Index, median (IQR)	n/a	3 (2,4)
COPS2, median (IQR)	n/a	46 (19,70)
Sepsis, <i>n</i> (%)	11,602 (12.5)	10,872 (12.0)
Acute organ failure type, <i>n</i> (%)		
Hematologic	8,544 (9.2)	2,260 (2.5)
Hepatic	1,658 (1.8)	905 (1.0)
Metabolic	3,759 (4.0)	4,319 (4.8)
Neurologic	6,741 (7.3)	3,459 (3.8)
Renal	11,669 (12.6)	14,716 (16.2)
Respiratory	27,288 (29.4)	12,145 (13.4)
Hospital LOS (days), mean (SD)	12 (14.5)	9.8 (9.9)
30-day mortality, <i>n</i> (%)	11,721 (12.6)	15,995 (17.6)

^a $p < 0.05$ for comparisons of all patient-level factors between the original and present study cohorts

SD, standard deviation; IQR, interquartile range; COPS2, COMorbidity Point Score; LOS, length of stay

Table 2

Discrimination of Acute Organ Failure Score (AOFS) claims-based model to predict 30-day mortality in the validation cohort (N=18,147)

Model ^a	c-Statistic	95% Confidence Interval
Original AOFS	0.644	0.639–0.649
Modified AOFS	0.721	0.711–0.730
Modified AOFS with COPS2	0.737	0.728–0.747
Modified AOFS with Elixhauser	0.748	0.739–0.757

^aThe Original AOFS used the β coefficients from the published model [11]; the Modified AOFS included the same variables as the Original but used β coefficients derived from the Medicare cohort; the COPS2 and Elixhauser index replaced the Charlson Comorbidity index in the modified AOFS

Table 3

Impact of additional comorbidity data on prediction of 30-day mortality in the validation cohort^a (*n* = 12,725)

Comorbidity Data Sources	Modified Acute Organ Failure Score				
	CCI Median (IQR) ^a	c-statistic (95% CI)	COPS2 Median (IQR) ^a	c-statistic (95% CI)	Elixhauser Median (IQR) ^a
Current hospital visit ^b	1 (0.3)	0.720 (0.708, 0.732)	46 (20.71)	0.737 (0.725, 0.748)	3 (2.4)
+ 1-year lookback at hospital visits ^{b,c}	2, (1.3)	0.726 (0.714, 0.738)	56 (28.90)	0.741 (0.729, 0.752)	3, (2.5)
+ 1-year lookback at hospital ^{b,c} + outpatient visits	4 (2.6)	0.728 (0.717, 0.740)	89 (49.140)	0.739 (0.728, 750)	4 (6.9)

^aExcluded admissions (29.9%) without associated provider-based inpatient claims in the Medicare Carrier file

^bIncluded only comorbidities designated as present on admission

^c16,807 (96.3%) of patients had no previous hospital visits containing an ICU visit in the prior year

IQR, interquartile range; CI confidence interval; CCI, Charlson Comorbidity Index; COPS2 COMorbidity Point Score 2