

ORIGINAL ARTICLE

Severity of Disease Estimation and Risk-Adjustment for Comparison of Outcomes in Mechanically Ventilated Patients Using Electronic Routine Care Data

Maaïke S. M. van Mourik, MD, PhD;^{1,2} Karel G. M. Moons, PhD;³ MICU Registry;⁴
 Michael V. Murphy, BA;¹ Marc J. M. Bonten, MD, PhD;^{2,3,*} Michael Klompas, MD, MPH^{1,4,*} for the Centers for Disease Control and Prevention (CDC) Prevention Epicenters

BACKGROUND. Valid comparison between hospitals for benchmarking or pay-for-performance incentives requires accurate correction for underlying disease severity (case-mix). However, existing models are either very simplistic or require extensive manual data collection.

OBJECTIVE. To develop a disease severity prediction model based solely on data routinely available in electronic health records for risk-adjustment in mechanically ventilated patients.

DESIGN. Retrospective cohort study.

PARTICIPANTS. Mechanically ventilated patients from a single tertiary medical center (2006–2012).

METHODS. Predictors were extracted from electronic data repositories (demographic characteristics, laboratory tests, medications, microbiology results, procedure codes, and comorbidities) and assessed for feasibility and generalizability of data collection. Models for in-hospital mortality of increasing complexity were built using logistic regression. Estimated disease severity from these models was linked to rates of ventilator-associated events.

RESULTS. A total of 20,028 patients were initiated on mechanical ventilation, of whom 3,027 deceased in hospital. For models of incremental complexity, area under the receiver operating characteristic curve ranged from 0.83 to 0.88. A simple model including demographic characteristics, type of intensive care unit, time to intubation, blood culture sampling, 8 common laboratory tests, and surgical status achieved an area under the receiver operating characteristic curve of 0.87 (95% CI, 0.86–0.88) with adequate calibration. The estimated disease severity was associated with occurrence of ventilator-associated events.

CONCLUSIONS. Accurate estimation of disease severity in ventilated patients using electronic, routine care data was feasible using simple models. These estimates may be useful for risk-adjustment in ventilated patients. Additional research is necessary to validate and refine these models.

Infect. Control Hosp. Epidemiol. 2015;36(7):807–815

Evaluating quality of care has become increasingly important over the past years and mandatory public reporting and non-payment programs for medical complications have strengthened the need for reliable and comparable measures of performance.^{1,2} Reporting mandates have been introduced for healthcare-associated infections in a majority of the United States as well as other countries, and nonpayment regulations are expanding.^{3–5} The emphasis on measuring quality of care by patient outcomes requires not only reliable ascertainment

of clinical outcomes but also accurate and practical methods of adjustment for underlying disease severity so as to allow for valid comparisons between healthcare facilities.^{1,6–10}

The ventilator-associated event (VAE) paradigm for surveillance of complications of mechanical ventilation (MV) was introduced as an alternative to traditional assessment of ventilator-associated pneumonia.¹¹ Two tiers of the new framework—ventilator-associated condition (VAC) and infection-related VAC (IVAC)—are being considered for future use in

Affiliations: 1. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts; 2. Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands; 3. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 4. Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

*Shared last author.

Presented in part: IDWeek 2013; San Francisco, California; October 4, 2013 (Abstract p1076).

Received November 7, 2014; accepted March 14, 2015; electronically published April 17, 2015

© 2015 by The Society for Healthcare Epidemiology of America. All rights reserved. 0899-823X/2015/3607-0008. DOI: 10.1017/ice.2015.75

benchmarking and pay-for-performance programs and are reportable to the Centers for Disease Control and Prevention's National Healthcare Safety Network (NHSN).¹¹ However, case-mix adjustment in NHSN is currently limited to stratification by type of intensive care unit (ICU) and better methods of incorporating disease severity are needed to facilitate more meaningful comparison of rates.^{12,13}

Several ICU mortality risk scoring systems, such as the Acute Physiology and Chronic Health Evaluation (APACHE) and Simplified Acute Physiology Score, have been in use for numerous years to measure disease severity in research and outcome evaluations.^{8,14–16} Although these models have good predictive accuracy for in-hospital mortality, their universal adoption is hampered by their inclusion of many clinical parameters that are labor-intensive to gather and infrequently available in a structured electronic format.^{17,18} Increasing use of electronic health records (EHRs) has fueled interest in the possibility of using routine care electronic data as a reliable alternative to time-consuming manual data collection. Although some risk scoring systems have been adapted to electronic data, their generalizability is questionable since they use diagnosis codes from administrative data, include clinical variables that are available in only a fraction of EHRs, and incorporate some variables that are not routinely measured in all patients.^{19–21} Furthermore, existing disease severity models may not be optimal for VAE risk-adjustment since they were derived from both ventilated and nonventilated patients and in some cases excluded cardiovascular surgery patients.^{14–16,20}

In this study, we aimed to develop a disease severity prediction model targeted specifically at mechanically ventilated patients using routinely collected, electronically stored data. We strove to optimize the balance between model precision, complexity, and generalizability. In addition, we assessed the relation between estimated disease severity and the risk of VAEs.

METHODS

Study Population

All patients initiated on MV, through an endotracheal tube or tracheostomy, and requiring ICU care in the Brigham and Women's Hospital (Boston, MA) from January 1, 2006, through December 31, 2012, were eligible for study inclusion. An interruption of MV of more than 1 calendar day defined the start of a new ventilation episode. If multiple hospital admissions or MV episodes occurred for a patient within the study period, a random hospitalization was chosen and the first episode within the hospital admission was selected. Patients younger than 18 years, those who could not be successfully linked to electronic data repositories, and those who were not admitted to an ICU were excluded. This study was reviewed by the institutional review board of Brigham and Women's Hospital and the requirement for individual patient consent was waived (protocol number: 2012P002052).

Outcomes

The primary outcome in developing a severity of disease prediction model was in-hospital mortality. As a sensitivity analysis, models were also developed for mortality within 30 days of initiation of MV as determined from hospital vital statistics records.^{22–24}

VACs and IVACs were selected as targets for risk-adjustment. A VAC reflects a sustained period of worsening oxygenation requiring increased ventilator settings whereas an IVAC is defined as a VAC with concurrent initiation of new antimicrobial therapy in combination with signs of infection (abnormal temperature, white blood cell count). Both were operationalized electronically in accordance with January 2013 NHSN specifications using ventilator settings derived from the respiratory therapy department.¹²

Predictors

Potential predictors were collected from the hospital's electronic Research Patient Data Registry.²⁵ This database compiles a broad variety of structured electronic data elements recorded during routine clinical care. We extracted data on patient demographic characteristics, laboratory tests, microbiology results, procedure codes, diagnosis codes, billing information, and medication use. ICU admission and discharge dates, however, were not available. Variables were selected for extraction on the basis of clinical associations with severity of disease and previously published models.^{14,15,20,26}

Subsequently, all candidate variables were assessed for potential inclusion. A qualitative estimate of feasibility was based on the extent of missing data and the expected likelihood that similar data will be available in electronic form in other hospitals. We also evaluated candidate variables for generalizability based on the probability that predictors would have similar interpretation when applied in other settings or across time (Table 1). When applicable, predictors were classified in broad groups that are expected to translate to other settings. ICU types were grouped into medical, surgical, neuroscience, and cardiovascular surgery categories, and procedure types were classified by anatomic or functional location (eg, musculoskeletal, cranial, digestive tract).

We collected values for predictors from the calendar day of MV onset and the day prior when available. Data were cleaned and processed as specified in sections 1A–1C of the online-only material. For each predictor, the observation reflecting the worst prognosis was taken (eg, lowest hematocrit, highest creatinine). Comorbid conditions were assessed based on administrative data using the Elixhauser method.^{27,28}

Model Development and Statistical Methods

Patterns of missing data were explored and multiple imputation was performed (mice package, version 10, R).²⁹ Exploratory

TABLE 1. Overview of Electronic Data Sources

Predictor category	Description	Expected feasibility ^a	Expected generalizability ^b	% Missing (comments)
Demographic characteristics	Age, sex	+	+	None
Unit type ^c	Medical, surgical, neuroscience, or cardiovascular surgery	+	+	None
Time to ventilator	Days from hospital admission to start of MV	+	+	None
Microbiology	Any blood culture performed (regardless of result)	+	+	None
Procedure codes	Classified in broad categories: musculoskeletal, thorax, CABG, valve, heart/lung transplant, aneurysm, digestive tract, spleen, cranial	+	+	None
Physiology I	Hematocrit, white blood cell count, platelets, creatinine, glucose, sodium, INR, blood urea nitrogen	+	+	1.2%
Physiology II	Albumin, total bilirubin, ALT	+/-	+/-	40.0%
Physiology III	Blood gas analysis (pH, pO ₂ , pCO ₂)	+/-	+/-	6.9%
Medication	(Groups of) medications initiated: Antibiotic (2 or more), antifungal, amiodarone, number of vasopressors, anticonvulsant, steroid, vitamin K, airway medication, antipsychotic	+/-	-	None (dependent on clinical practice)
Comorbidities	Elixhauser comorbidity groups (based on ICD9 and DRG codes)	+/-	-	None (dependent on coding practices)

NOTE. Data were collected during the process of routine care on the first calendar day of mechanical ventilation (MV) or the day prior when available. All groups of predictors are classified by anticipated feasibility and generalizability of electronic data collection. For details on variable definitions please refer to supplementary material (section 1). ALT, alanine transaminase; CABG, coronary artery bypass graft; DRG, diagnosis-related group; ICD9, *International Classification of Disease, Ninth Edition*; INR, international normalized ratio.

^aFeasibility was estimated on the basis of the amount of missing data in our sample and the expected availability from electronic health records in other settings.

^bGeneralizability was defined as the anticipated consistency of predictor interpretation in other settings. +: good anticipated feasibility/generalizability, +/-: some concerns, -: limited anticipated feasibility/generalizability.

^cSurgical intensive care unit (ICU): all surgical ICU types (eg, general, trauma, thoracic); Neuroscience ICU: neurology and neurosurgery.

univariable analysis of the association between predictors and in-hospital mortality was performed and restricted cubic splines were used for continuous variables that did not have a linear association with the log-odds of mortality (online-only material, section 1B).³⁰

Regression models of incremental complexity were built, first including groups of candidate predictors with the best expected feasibility and generalizability (defined in Table 1). All variables within each group of predictors were added to the model irrespective of their individual significance. Interactions detected by backward elimination were judged relevant if $P < .0001$; very rare combinations of predictors were excluded. If a group of predictors no longer contributed to model performance, measured by the Akaike information criterion, it was eliminated in its entirety. To limit overfitting and to obtain bias-corrected estimates of model performance, internal validation and uniform shrinkage were performed using bootstrap samples (100 samples per imputation). This method generates more accurate estimates of performance than split-sample or cross-validation approaches.³¹ Parameter estimates from the imputation sets were pooled by Rubin's rule.³²

Model discrimination and calibration were assessed using the area under the receiver operating characteristic curve (AUROC), the Brier score, the Hosmer-Lemeshow test, and examination of calibration plots. Given the large heterogeneity in mortality across ICU types, the model with the most

favorable balance of complexity to performance was developed stratified by ICU type.

Finally, the possibility of VAE risk-adjustment using the severity of underlying disease estimates was assessed by examining the relation between predicted risk of mortality and observed incidence of VAEs. For the overall population and for each type of ICU separately, patients were classified by decile of predicted mortality risk and the incidence of VACs and IVACs was determined within each decile. In addition, attention was paid to the relation between disease severity and the competing outcomes for VAEs. Because patients must by definition have received at least 4 days of MV—2 days of stability or improvement followed by 2 days of deterioration—in order to be at risk of VAE, early death or early discharge within 4 days can be considered competing outcomes.³³

All analyses were performed using SAS, version 9.2 (SAS Institute), and R (www.r-project.org).

RESULTS

During the study period, 24,366 episodes of MV were recorded in 20,581 unique patients. Of these patients, 59 were younger than 18 years; 350 did not have a successful medical record match; and 144 were not admitted to an ICU, thus leaving 20,028 patients for analysis. Median age was 64 years, 60.3% were male, and 40.7% were admitted to the cardiovascular

surgery unit (Table 2). Average in-hospital mortality was 15.1%, ranging from 3.3% for cardiovascular surgery to 38% for the medical ICUs.

As expected, models using larger numbers of predictors had improved performance characteristics, although the gains with incrementally more complicated models became smaller (Table 3).

TABLE 2. Selected Patient Characteristics in Comparison Study of Outcomes in Mechanically Ventilated Patients Using Electronic Routine Care Data, 2006–2012

Variable	All episodes n = 20,028	Discharged alive n = 17,001	Deceased in hospital n = 3,027
Age, median (IQR), y	64 (53–73)	64 (52–74)	67 (56–78)
Male sex	12,084 (60.3)	10,362 (60.9)	1,722 (56.9)
Unit type			
Medical	4,282 (21.4)	2,663 (15.7)	1,619 (53.5)
Surgical	5,347 (26.7)	4,720 (27.8)	627 (20.7)
Neuroscience	2,252 (11.2)	1,742 (10.2)	510 (16.8)
Cardiovascular surgery	8,147 (40.7)	7,876 (46.3)	271 (9.0)
Time to ventilation onset			
Day of hospital admission	8,507 (42.5)	7,333 (43.1)	1,174 (37.8)
Day 1 or 2 after admission	7,097 (35.4)	6,257 (36.8)	840 (27.8)
Days 3–7 after admission	3,017 (15.1)	2,479 (14.6)	538 (17.8)
≥8 days after admission	1,407 (7.0)	932 (5.5)	475 (15.7)
Any surgical procedure ^a	13,457 (67.2)	12,949 (76.2)	963 (31.8)
Blood culture specimen obtained ^a	4,315 (21.5)	2,860 (16.8)	1,455 (48.1)
Duration of MV, median (IQR), d	2 (1–3)	2 (1–3)	3 (2–8)
Length of hospital stay, median (IQR), d	11 (7–18)	11 (8–18)	9 (4–19)

NOTE. Data are no. (%) unless otherwise indicated. IQR, interquartile range; MV, mechanical ventilation.

^aOn day of MV onset or day prior.

TABLE 3. Results of Model Development and Internal Validation for In-Hospital Mortality

Predictor category	Model 1	Model 2	Model 3	Model 4	Model 5
Demographic characteristics	x	x	x	x	x
Unit type	x	x	x	x	x
Time to ventilation	x	x	x	x	x
Microbiology ^a	x	x			
Procedure codes	x	x	x	x	x
Physiology I ^b		x	x	x	x
Physiology II ^b + III ^b			x	x	x
Medication				x	x
Comorbidities					x
N independent predictors	17	36	47	56	85
AUROC ^c	0.829	0.870	0.877	0.879	0.883
(95% CI)	(0.822–0.837)	(0.864–0.877)	(0.871–0.883)	(0.873–0.885)	(0.878–0.889)
Medical ICU	0.674	0.758	0.775	0.783	0.796
Surgical ICU	0.732	0.822	0.835	0.840	0.843
Neuroscience ICU	0.724	0.784	0.786	0.794	0.794
CV surgery	0.769	0.852	0.860	0.857	0.871
H-L statistic	12.2	50.8	59.5	49.9	55.1
P value H-L	.140	<.001	<.001	<.001	<.001
Brier score	0.102	0.091	0.089	0.088	0.087

NOTE. Models are ordered by increasing complexity as classified by feasibility and generalizability of data collection. AUROC, area under the receiver operating characteristic curve; CV, cardiovascular; H-L, Hosmer-Lemeshow; ICU, intensive care unit.

^aMicrobiology (blood culture performed) did not contribute to model performance in models 3 and up and was therefore removed.

^bPhysiology I-III, please refer to Table 1 and supplementary data for definitions.

^cBias-corrected.

Model 2, using only demographic information, unit type, microbiology data, procedure codes, and commonly performed laboratory tests, achieved an optimism-corrected AUROC of 0.870 (95% CI, 0.864–0.876) (parameter estimates in the online-only material Tables S1, S2). On internal validation, overoptimism was very limited (0.003–0.006 decrease in AUROC). Calibration was adequate with slight overprediction for high-risk patients (online-only material, Figure S1); this is also reflected by the Hosmer-Lemeshow test statistics (χ^2 test, 49.9 to 59.5 on 10 degrees of freedom). A sensitivity analysis for 30-day mortality demonstrated similar performance of all models (online-only material, Table S3).

Model 2 also had reasonable performance when applied within ICU types (AUROC, 0.758 to 0.852); as can be expected, the observed performance was somewhat lower because differences between patients within each of the unit types were smaller than across units. Given the large heterogeneity in mortality rates between types of ICUs and the intended application of achieving adequate estimation of disease severity within ICU types, the model was refit stratified by ICU type (Table 4, Table S4), resulting in similar performance yet less complex models because interaction terms were no longer necessary. In addition, the performance of this model was compared with APACHE II scores of a subset of 144 patients included in the hospital's Registry of Critical Illness who were mechanically ventilated on admission to the ICU.³⁴ This revealed an AUROC of 0.66 for APACHE II versus 0.80 for the final model.

Depending on the type of ICU, the overall risk of VAC ranged from 0.9% to 9.1% and of IVAC from 0.4% to 3.2% (Table 4). Overall, there was a positive association between the estimated severity of disease (predicted mortality from model 2 stratified by ICU type) and the occurrence of VAC and IVAC (Figure 1). Specifically, for surgical and cardiovascular surgery units there was a positive association between estimated disease severity and VAE incidence, whereas for medical and

neurological ICUs the picture was less clear. As can be seen in the bar charts, the association between observed risk of VAE and estimated disease severity is largely explained by the proportion of patients who remain on MV 4 days or longer and are thus at risk of developing a VAE. In particular for medical units, the decline in VAC incidence in higher mortality risk groups is caused by an increasing proportion of patients experiencing the competing event of dying within 4 days of MV onset.

DISCUSSION

Adjustment for severity of underlying disease is a necessary requirement to achieve valid comparisons of patient outcomes across healthcare providers. This study is meant as a proof-of-concept that estimating underlying disease severity in mechanically ventilated patients is feasible using fairly simple models applied to electronic routine care data. The performance characteristics obtained were in the same range as those observed in the derivation studies of more labor-intensive methods such as APACHE and Simplified Acute Physiology Score.⁸ Models of increasing complexity showed marginally better performance but at the cost of more extensive data collection and rising concerns with generalizability to different settings.

In addition, this study found an intricate relation between estimated disease severity and the incidence of VAEs, thus providing a starting point towards improved correction for differences in patient populations when these outcomes are used for assessment of quality of care. Currently, event rates reported to NHSN are stratified by unit type for comparison.¹³ Although unit type does explain an important fraction of the observed variation in VAE incidence, this can be further refined by stratifying on underlying disease severity within ICU types. This association between estimated disease severity

TABLE 4. Model Development for In-Hospital Mortality Stratified by Unit Type for Model 2

Variable	Medical	Surgical	Neuroscience	Cardiovascular surgery
MV episodes	4,282	5,347	2,252	8,147
Deceased, N (%)	1,619 (37.8)	627 (11.7)	510 (22.6)	271 (3.3)
Mortality				
AUROC ^a	0.753	0.818	0.770	0.856
(95 % CI)	(0.738–0.767)	(0.802–0.835)	(0.749–0.792)	(0.835–0.877)
H-L statistic	10.9	21.9	6.0	12.9
Brier score	0.190	0.084	0.142	0.027
Ventilator-associated events				
N VAC (/100 MV episodes)	390 (9.1)	402 (7.5)	112 (5.0)	67 (0.9)
N IVAC (/100 MV episodes)	135 (3.2)	167 (3.1)	51 (2.3)	33 (0.4)
VAC rate/1,000 MV days	16.1	15.6	10.4	4.3
IVAC rate/1,000 MV days	5.6	6.5	4.8	2.1

NOTE. AUROC, area under the receiver operating characteristic curve; H-L, Hosmer-Lemeshow; IVAC, infection-related ventilator-associated condition; MV, mechanical ventilation; VAC, ventilator-associated condition.

^aFor mortality, after correction for overoptimism. Models were developed by unit type; procedure types that did not occur within the unit types were excluded. There were no significant interactions in all models.

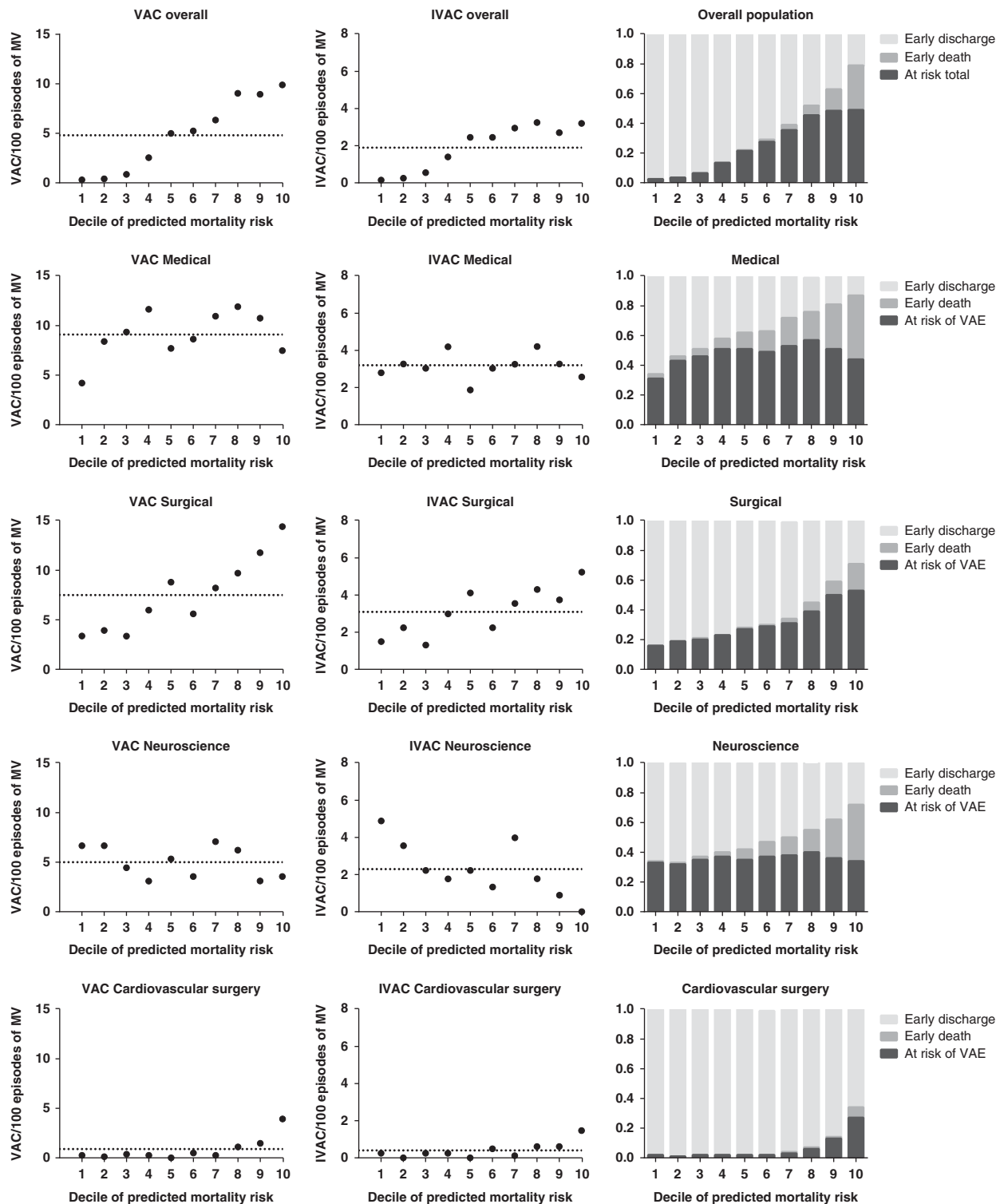


FIGURE 1. Association between estimated disease severity (predicted risk of in-hospital mortality) and the occurrence of ventilator-associated events (VAEs). Predicted risk of in-hospital mortality was derived from the final model stratified by unit type and incidence of ventilator-associated condition (VAC, left column) and infection-related VAC (IVAC, middle column). Horizontal lines represent the overall event incidence. The bar charts (right column) depict the proportion of patients at risk of developing a VAC or IVAC and those either discharged alive off the ventilator within 4 days of ventilator onset (early discharge) or deceased within 4 days of ventilator onset (early death). Top row: deciles of predicted mortality risk calculated for the overall population; bottom 4 rows: patients are stratified by decile of mortality risk within each intensive care unit type. MV, mechanical ventilation.

and VAE risk seems to be mediated by identifying the proportion of patients who remain on MV long enough to be eligible for developing a VAE ($MV \geq 4$ days). We purposely included all patients initiated on MV regardless of whether they achieved 4 days of MV because all MV days are counted in the denominator of VAE rates and the duration of MV is also unknown at the time of ventilation onset (the moment of prediction). Future research is needed to determine how to assess the exact relationship between severity of underlying disease and the likelihood of developing VAE and how to account for the competing outcomes of early death and early discharge. Because the models were developed in a single tertiary care US hospital, further validation studies are needed to determine their performance and utility in different hospitals and types of ICU. These studies will also provide an opportunity to refine estimates for a broader range of (specialized) ICU types, including mixed medical-surgical units and units affiliated with smaller community hospitals, ideally in alignment with the ICU classifications currently employed by NHSN.

Although numerous models have previously been developed to predict in-hospital mortality, these were not developed for the purpose of comparing patient outcomes in the specific domain of mechanically ventilated patients, and several concerns limit their application in risk-adjustment for VAEs, including the highly divergent degrees of sophistication in EHR contents across hospitals.^{14,15,26} Existing models, such as the APACHE, require extensive amounts of manual data collection, including physiologic variables that are not routinely documented in EHRs (eg, Glasgow coma scale, blood pressure, urine output) and reason for admission, which is subject to considerable interrater variability.^{18,35} Strikingly, models in some cases included administrative data collected after the first 2 days of admission, which may better reflect the evolution of disease within hospitals rather than patients' presenting conditions.^{20,21} Moreover, translating a model directly to an electronic environment may introduce unanticipated loss of comparability.^{21,36} The models developed in this study aim to address these concerns and do have similarities with existing risk-adjustment models, in particular the model developed within the Veteran's Affairs system.²⁰ However, that model makes extensive use of admission diagnoses and procedure codes and includes more laboratory variables than the simpler models developed in our study.

The process of model development in this study was targeted specifically for use with electronic data and an important component was the a priori assessment of the feasibility and generalizability of data collection through EHRs. Preference was given to measuring information "close to the patient" as opposed to proxies such as administrative coding data. We found that simple models without admission diagnoses, disease codes, or medication use—variables for which the use and classification are more likely to differ across settings—were still able to perform robustly. We further incorporated methods for handling of outliers and approaches to handling of missing observations other than imputing a normal value since the latter weakens the reliability of risk-adjustment.³⁷

Several important additional limitations need to be addressed. Firstly, this study was developed in a single tertiary medical center using data from a 6-year period and did not include a separate validation sample; use of a random split-sample for validation was judged not to provide sufficient new insights in comparison with the internal validation strategy used.³¹ As described above, validation studies will be needed before large-scale adoption. Nevertheless, the results of this study will inform the design and data collection of such studies. The Hosmer-Lemeshow statistics for our models suggested significant miscalibration but this test is known to be unreliable with large sample sizes and assessment of calibration plots showed adequate calibration.³⁸ Finally, examination of yearly observed-to-expected ratios by unit type did not reveal any secular decreases in model performance (online supplementary data).

We could perform only a very limited head-to-head comparison of our models with other, existing, methods (eg, APACHE or Simplified Acute Physiology Score) since these severity-of-disease estimators are not routinely collected in our setting; their absence was the driving force behind the current study. It has been argued that using 30-day mortality is more reliable than in-hospital mortality for the purpose of outcome comparison; however, data on in-hospital mortality is often collected more consistently than 30-day mortality.^{23,24} In this study, model performance characteristics were similar for both. Prior studies have identified several possible predictors that could not be collected from our EHRs but that may be of predictive value in future studies, including (hospital) location prior to ventilator onset, time to ICU admission, transfer status, urgency of admission and/or surgery, and the presence of do-not-resuscitate orders.^{14,15,20,39} In addition, inclusion of ICU type as a predictor requires assumptions to be made regarding similarity of patient characteristics across ICU types in different hospitals. Future validation studies will need to assess robustness to differences in admission policies and define methods to handle specialty ICUs. The use of more sophisticated modeling strategies such as supervised machine learning have been suggested in the recent literature as efficient methods of model development, and focusing on using parameters that can be collected reliably and efficiently would be of added value.¹⁹

Valid comparison of patient outcomes and occurrence of VAEs requires adequate methods of correction for severity of underlying disease. This study demonstrates the feasibility of estimating disease severity in ventilated patients using electronic routine care data and provides insight into the balance between complexity of data collection and model performance. The estimates of underlying disease severity obtained from these models show utility in future adjustment of VAE rates, although further research is necessary to validate and refine the proposed models and develop practical and robust methods of benchmarking.

ACKNOWLEDGMENTS

We thank the Brigham and Women's Hospital Registry of Critical Illness investigators for contributing APACHE II scores for our comparative

analysis: Carolina Quintana, MD, Rebecca M. Baron, MD, Laura E. Fredenburgh, MD, Joshua A. Englert, MD, Lee Gazourian, MD, Anthony F. Massaro, MD, Anna Trtchounian, Tamas Dolinay, MD, PhD, and Augustine M. K. Choi, MD. In addition, we thank N.P.A. Zuithoff, PhD, for his statistical advice.

Financial support. Netherlands Organization for Scientific Research (project 9120.8004 and 918.10.615 to K.G.M.M.); Centers for Disease Control and Prevention (grant 1U54CK000172 to M.K.).

Potential conflicts of interest. All authors report no conflicts of interest relevant to this article.

Address correspondence to Maaikje S. M. van Mourik, Department of Medical Microbiology, HP G04.614, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands (M.S.M.vanMourik-2@umcutrecht.nl).

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/ice.2015.75>

REFERENCES

- Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA* 2013;310:1971–1980.
- Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med* 2007;357:1573–1575.
- Centers for Medicare and Medicaid Services (CMS). Hospital-acquired conditions. CMS website. Available at: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html. Updated September 20, 2012. Accessed November 25, 2013.
- Association of Professionals in Infection Control and Epidemiology (APIC). HAI reporting laws and regulations. APIC website. Available at: http://www.apic.org/Resource_/TinyMceFileManager/Advocacy-PDFs/Advocacy_Updates/Static_map_-_HAI_revised_7-6-11.gif. Updated July 6, 2011. Accessed February 13, 2012.
- Haustein T, Gastmeier P, Holmes A, et al. Use of benchmarking and public reporting for infection control in four high-income countries. *Lancet Infect Dis* 2011;11:471–481.
- Mu Y, Edwards JR, Horan TC, Berrios-Torres SI, Fridkin SK. Improving risk-adjusted measures of surgical site infection for the National Healthcare Safety Network. *Infect Control Hosp Epidemiol* 2011;32:970–986.
- Harris AD, McGregor JC. The importance of case-mix adjustment for infection rates and the need for more research. *Infect Control Hosp Epidemiol* 2008;29:693–694.
- Breslow MJ, Badawi O. Severity scoring in the critically ill, part 1: interpretation and accuracy of outcome prediction scoring systems. *Chest* 2012;141:245–252.
- van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. *Br J Surg* 2013;100:628–636.
- Kritsotakis EI, Dimitriadis I, Roubelaki M, et al. Case-mix adjustment approach to benchmarking prevalence rates of nosocomial infection in hospitals in Cyprus and Greece. *Infect Control Hosp Epidemiol* 2008;29:685–692.
- Magill SS, Klompas M, Balk R, et al. Developing a new, national approach to surveillance for ventilator-associated events. *Crit Care Med* 2013;41:2467–2475.
- Centers for Disease Control and Prevention (CDC). Surveillance for ventilator-associated events. CDC website. Available at: <http://www.cdc.gov/nhsn/acute-care-hospital/vae/>. Updated November 25, 2013. Accessed November 26, 2013.
- Dudeck MA, Weiner LM, Ien-Bridson K, et al. National Healthcare Safety Network (NHSN) report, data summary for 2012, device-associated module. *Am J Infect Control* 2013;41:1148–1166.
- 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.
- Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006;34:1297–1310.
- Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). *Crit Care Med* 2007;35:827–835.
- Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in US hospitals. *N Engl J Med* 2009;360:1628–1638.
- 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.
- Johnson AE, Kramer AA, Clifford GD. A new severity of illness scale using a subset of Acute Physiology, Age, and Chronic Health Evaluation data elements shows comparable predictive accuracy. *Crit Care Med* 2013;41:1711–1718.
- Render ML, Deddens J, Freyberg R, et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. *Crit Care Med* 2008;36:1031–1042.
- Liu V, Turk BJ, Ragins AI, Kipnis P, Escobar GJ. An electronic Simplified Acute Physiology Score-based risk adjustment score for critical illness in an integrated healthcare system. *Crit Care Med* 2013;41:41–48.
- Graham PL, Cook DA. Prediction of risk of death using 30-day outcome: a practical end point for quality auditing in intensive care. *Chest* 2004;125:1458–1466.
- Drye EE, Normand SL, Wang Y, et al. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. *Ann Intern Med* 2012;156:19–26.
- Pouw ME, Peelen LM, Moons KG, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *BMJ* 2013;347:f5913.
- Partners Healthcare. Research Patient Data Registry (RPDR). Partners Healthcare website. Available at: <http://rc.partners.org/rpdr>. Updated November 1, 2013. Accessed November 11, 2013.
- Higgins TL, Kramer AA, Nathanson BH, Copes W, Stark M, Teres D. Prospective validation of the intensive care unit admission Mortality Probability Admission Model (MPM0-III). *Crit Care Med* 2009;37:1619–1623.
- Healthcare Cost and Utilization Project. Comorbidity Software, version 3.7. Available at: <http://www.hcup-us.ahrq.gov/tools-software/comorbidity/comorbidity.jsp>. Updated 2012. Accessed November 1, 2013.

28. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
29. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariable Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
30. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag, 2001.
31. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–781.
32. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken: J. Wiley & Sons, 1987.
33. Centers for Disease Control and Prevention (CDC) Division of Healthcare Quality Promotion. *The National Healthcare Safety Network Device-Associated Module: Ventilator-Associated Event Protocol*. Atlanta: CDC, 2013.
34. Dolinay T, Kim YS, Howrylak J, et al. Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med* 2012;185:1225–1234.
35. Polderman KH, Thijs LG, Girbes AR. Interobserver variability in the use of APACHE II scores. *Lancet* 1999;353:380.
36. Bosman RJ, Oudemane van Straaten HM, Zandstra DF. The use of intensive care information systems alters outcome prediction. *Intensive Care Med* 1998;24:953–958.
37. Afessa B, Keegan MT, Gajic O, Hubmayr RD, Peters SG. The influence of missing components of the Acute Physiology Score of APACHE III on the measurement of ICU performance. *Intensive Care Med* 2005;31:1537–1543.
38. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med* 2007;35:2052–2056.
39. Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0-III and influence of resuscitation status on model performance. *Chest* 2012;142:851–858.