

OPEN

# Risk Adjustment for Sepsis Mortality to Facilitate Hospital Comparisons Using Centers for Disease Control and Prevention's Adult Sepsis Event Criteria and Routine Electronic Clinical Data

Chanu Rhee, MD, MPH<sup>1,2</sup>; Rui Wang, PhD<sup>1,3</sup>; Yue Song, MSc<sup>1,3</sup>; Zilu Zhang, MS<sup>1,4</sup>; Sameer S. Kadri, MD MSc<sup>5</sup>; Edward J. Septimus, MD<sup>1,6</sup>; David Fram, BA<sup>7</sup>; Robert Jin, MS<sup>1</sup>; Russell E. Poland, PhD<sup>1,8</sup>; Jason Hickok, MBA<sup>9</sup>; Kenneth Sands, MD, MPH<sup>1,8</sup>; Michael Klompas, MD, MPH<sup>1,2</sup> for the Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program

<sup>1</sup>Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Boston, MA.

<sup>2</sup>Department of Medicine, Brigham and Women's Hospital, Boston, MA.

<sup>3</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

<sup>4</sup>Department of Medical Oncology, Harvard Medical School/Dana Farber Cancer Institute, Boston, MA.

<sup>5</sup>Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD.

<sup>6</sup>Department of Internal Medicine, Texas A&M College of Medicine, Houston, TX.

<sup>7</sup>Commonwealth Informatics, Waltham, MA.

<sup>8</sup>Clinical Services Group, HCA Healthcare, Nashville, TN.

<sup>9</sup>Ondine Biomedical, Vancouver, BC, Canada.

Supplemental digital content is available for this article. Direct URL citations appear in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

Supported, in part, by grants from the Centers for Disease Control and Prevention (U54CK000484), Agency for Healthcare Research and Quality (K08HS025008 to Dr. Rhee), and intramural funds from the National Institutes of Health Clinical Center and National Institute of Allergy and Infectious Diseases (Dr. Kadri).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Centers for Disease Control and Prevention, Agency for Healthcare Research and Quality or the National Institutes of Health.

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: [crhee@bwh.harvard.edu](mailto:crhee@bwh.harvard.edu)

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

*Crit Care Expl* 2019;1:e0049

DOI: 10.1097/CCE.0000000000000049

**Objectives:** Variability in hospital-level sepsis mortality rates may be due to differences in case mix, quality of care, or diagnosis and coding practices. Centers for Disease Control and Prevention's Adult Sepsis Event definition could facilitate objective comparisons of sepsis mortality rates between hospitals but requires rigorous risk-adjustment tools. We developed risk-adjustment models for Adult Sepsis Events using administrative and electronic health record data.

**Design:** Retrospective cohort study.

**Setting:** One hundred thirty-six U.S. hospitals in Cerner HealthFacts (derivation dataset) and 137 HCA Healthcare hospitals (validation dataset).

**Patients:** A total of 95,154 hospitalized adult patients (derivation) and 201,997 patients (validation) meeting Centers for Disease Control and Prevention Adult Sepsis Event criteria.

**Interventions:** None.

**Measurements and Main Results:** We created logistic regression models of increasing complexity using administrative and electronic health record data to predict in-hospital mortality. An administrative model using demographics, comorbidities, and coded markers of severity of illness at admission achieved an area under the receiver operating curve of 0.776 (95% CI, 0.770–0.783) in the Cerner cohort, with diminishing calibration at higher baseline risk deciles. An electronic health record–based model that integrated administrative data with laboratory results, vasopressors, and mechanical ventilation achieved an area under the receiver operating curve of 0.826 (95% CI, 0.820–0.831) in the derivation cohort and 0.827 (95% CI, 0.824–0.829) in the validation cohort, with better calibration than the administrative model. Adding vital signs and Glasgow Coma Score minimally improved performance.

**Conclusions:** Models incorporating electronic health record data accurately predict hospital mortality for patients with Adult Sepsis Events and outperform models using administrative data alone. Utilizing laboratory test results, vasopressors, and mechanical ventilation without vital signs may achieve a good balance between data collection needs and model performance, but electronic health record-based models must be attentive to potential variability in data quality and availability. With ongoing testing and refinement of these risk-adjustment models, Adult Sepsis Event surveillance may enable more meaningful comparisons of hospital sepsis outcomes and provide an important window into quality of care.

**Key Words:** adult sepsis event; electronic health records; hospital benchmarking; risk adjustment; sepsis; surveillance

There is intense interest in improving and measuring the quality of sepsis care provided by U.S. hospitals (1). Current benchmarking efforts, such as the Centers for Medicare and Medicaid Services Severe Sepsis and Septic Shock Early Management Bundle ("SEP-1"), primarily focus on compliance with time-sensitive process measures rather than sepsis outcomes (2). The development of a reliable risk-adjusted sepsis outcome measure could complement process metrics and provide a broader window into variations in quality of care (3).

Previous studies have documented wide variability in hospital risk-adjusted sepsis mortality rates using administrative data (4–6), but these are potentially confounded by large variations in diagnosis, documentation, and coding practices for sepsis and organ dysfunction that markedly affect perceived levels of severity of illness (7). Surveillance using electronic health record (EHR) data allows for more consistent case finding while still allowing for the possibility of automated national surveillance.

In March 2018, the U.S. Centers for Disease Control and Prevention (CDC) released the "Adult Sepsis Event" surveillance definition that uses objective clinical data routinely available in EHRs to identify sepsis (8). Compared with the Sequential Organ Failure Assessment score used in Sepsis-3 criteria, the Adult Sepsis Event definition identifies similar types of patients, has comparable or better predictive validity for mortality, and may facilitate more widespread automated surveillance due to its simpler organ dysfunction criteria (9). Although primarily meant to help hospitals track internal sepsis rates and outcomes, a standardized surveillance definition could serve as the foundation for credible hospital comparisons if coupled with rigorous risk adjustment (10).

Sepsis risk-adjustment models using administrative data already exist (4, 11, 12), sometimes supplemented with manually collated clinical data (e.g., the Surviving Sepsis Campaign database, mandatory reporting requirements in New York state) (13, 14). There are also validated mortality prediction scores for critically ill patients, including the Acute Physiology and Chronic Health Evaluation (APACHE) and Simplified Acute Physiology Score (SAPS) scores (15–17). These models, however, may not be ideal for routine operational benchmarking of sepsis outcomes because they either have limited granularity and variable accuracy (e.g., administrative data), require clinical parameters infrequently available in structured electronic format (e.g., urine

output, mental status, chills with rigors), or are derived from one geographic location (e.g., New York state). Scores specific to the ICU also may not suffice because more than 50% of sepsis patients never require critical care services (18, 19).

Our objective was to develop and validate risk-adjustment models for Adult Sepsis Events in two large cohorts of U.S. hospitals using EHR and administrative data. In doing so, we examined the incremental benefit of successive sets of covariates to understand the balance between data collection burden and model performance.

## METHODS

### Data Sources

This study was approved by the Institutional Review Board at Harvard Pilgrim Health Care Institute. Our dataset for model development was Cerner HealthFacts, a deidentified database that contains detailed clinical data from U.S. academic and community hospitals that use the Cerner EHR system (Cerner Corporation, Kansas City, MO) (9, 20–24). We included all adults ( $\geq 20$  yr old) admitted to 136 hospitals from January 2009 to September 2015 (20). We externally validated all models using 2013–2014 data from 137 hospitals in the HCA Healthcare network (20). HCA Healthcare includes urban, suburban, and rural community medical centers across 20 states and primarily use the Meditech EHR system (Medical Information Technology Incorporated, Westwood, MA) (25). Comprehensive clinical data are stored centrally and undergo line-item validation until more than 99% accuracy is achieved (20). Study hospital characteristics are summarized in **Supplemental Table 1** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>). Encounters with missing discharge dispositions and with *International Classification of Diseases*, 10th Revision, Clinical Modification codes instead of *International Classification of Diseases*, 9th Revision, Clinical Modification (ICD-9-CM) codes were excluded.

### Outcomes and Predictors

The primary outcome for our models was in-hospital mortality among patients meeting Adult Sepsis Event criteria during hospitalization. This definition flags "presumed serious infection" (blood culture order and new antibiotics continued for  $\geq 4$  d or until  $\leq 1$  d prior to death, discharge to hospice, or transfer to another hospital) and "concurrent organ dysfunction" (initiation of vasopressors or mechanical ventilation, elevated lactate, increase in baseline creatinine or total bilirubin, or decrease in baseline platelets) (for complete CDC criteria, see **Supplemental Table 2**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>) (8).

Potential predictors from administrative and EHR data included demographics, admission source, comorbidities, infectious syndromes, laboratory tests, medications, mechanical ventilation initiation, blood culture results, vital signs, and Glasgow Coma Scale (GCS) measurements (**Table 1**). For physiologic variables (laboratory tests, vital signs, GCS), we used the most abnormal values within  $\pm 1$  calendar day from the "day of sepsis onset," defined as the earliest day the blood culture or first qualifying antibiotic day occurred (concurrent with organ dysfunction). Elixhauser comorbidities (26) were supplemented with other comorbidities (based on ICD-9-CM

**TABLE 1. Overview of Predictors and Relevant Surveillance Characteristics**

Predictor Category	Variables	Expected Availability <sup>a</sup>	Objectivity <sup>a</sup>
Demographics	Age, sex, race	+++	+++
Comorbidities	Elixhauser comorbidity groups, leukemia, stem cell transplant, solid organ transplant (based on ICD-9-CM codes)	+++	+ (Dependent on coding practices)
Infection site	Pneumonia, urinary, intra-abdominal, skin/soft tissue, septicemia/bacteremia, obstetric/gynecologic, CNS, 2 or more of the above, unknown/none of the above (based on ICD-9-CM codes)	+++	+ (Dependent on coding practices)
Time to sepsis	Days from hospital admission to sepsis onset	++	+++
ICU at sepsis onset	Whether patient was in ICU on the day of sepsis onset	++	++ (Dependent on hospital criteria for ICU admission)
Adult Sepsis Event organ dysfunction variables	Number of vasopressors, mechanical ventilation, lactate, creatinine, bilirubin, platelet count	++	+++
Extended laboratory values	WBC, hematocrit, sodium, anion gap, albumin, aspartate aminotransferase, international normalized ratio	++	+++
Microbiology	Positive blood culture	+	+++
Vital signs	Systolic blood pressure, temperature, respiratory rate	+	+++
GCS <sup>b</sup>	GCS	+	++

GCS = Glasgow Coma Scale, ICD-9-CM = *International Classification of Diseases*, 9th Revision, Clinical Modification.

<sup>a</sup>“Expected availability” (the degree to which covariates would be available in different electronic health record systems) and “objectivity” (the resistance of covariates to differences in measurement or application between clinicians and hospitals) are rated qualitatively by 1, 2, or 3 “+” signs, with more “+” signs indicating stronger attributes in that category.

<sup>b</sup>Vital signs and GCS were unavailable in the HCA Healthcare dataset.

codes) that might influence sepsis mortality: leukemia (204–208), stem cell transplant (V42.81 and V42.82), and solid organ transplant (V42.0–42.7, V42.83, V42.84, V42.89, V42.9). ICD-9-CM codes for infectious syndromes were adapted from prior work (27, 28).

We rated the qualities of each potential predictor by “expected availability” (i.e., ubiquity in different EHR systems) and “objectivity” (i.e., consistent ascertainment between clinicians and hospitals) (Table 1). Administrative data, for example, should be ubiquitous, whereas mental status is not routinely recorded in all EHR systems. Basic vital signs and laboratory data are generally objective, but diagnosis and coding practices for infections and organ dysfunction differ between hospitals (7).

### Statistical Methods

We performed exploratory univariate analyses of the associations between each potential predictor and in-hospital mortality. For continuous physiologic variables that did not appear to have a linear association with the log odds of mortality, we included quadratic terms in the model. We then constructed regression models of incremental complexity. We first used administrative candidate predictors (“basic administrative model”): demographics, comorbidities, admission source, infectious diagnoses, ICU admission, and days from hospital admission to sepsis onset. We then implemented an “advanced administrative model” developed by Ford et al (11) that incorporates demographics, comorbidities, and proxies for severity of illness at admission (codes for mechanical ventilation, shock, hemodialysis, and ICU admission) (11).

We then added clinical data in a sequential fashion based on their expected availability. For the simplest model (clinical model 1), we only incorporated the data elements required to identify Adult Sepsis Events (vasopressors, mechanical ventilation, lactate, creatinine, bilirubin, and platelet counts on the day of sepsis onset). Subsequent models included additional laboratory data from chemistries, complete blood cell counts, and liver function tests (clinical model 2), and vital signs and GCS measurements (clinical model 3). Quadratic terms for continuous clinical variables were added to the model, and backward elimination was performed at a significance level of 0.05. To keep the model hierarchically well formulated, the linear term was retained in the model if the corresponding quadratic term was significant. To properly quantify the uncertainty in parameter estimates accounting for the clustering effect within hospitals, we used a sandwich variance estimator robust to misspecification of the working correlation structure (29).

All models were developed using two thirds of the Cerner cohort and internally validated on the remaining third. The models were then applied to the HCA Healthcare dataset for external validation using the coefficients developed with Cerner, with the exception of clinical model 3 as vital signs and GCS measurements were unavailable from HCA.

We handled missing severity-of-illness covariates in two ways in Cerner: multiple imputation (**supplemental methods**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>) and imputation of normal values (for normal values, see

**Supplemental Table 3**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>). We compared results from multiple imputation versus normal imputation, as is commonly done with clinical severity-of-illness scores (30–32). We reasoned that if successful, normal imputation could facilitate real-world application of our risk-adjustment models in facilities where more complicated strategies for handling missing data might be impractical. Performance of clinical models 1 and 2 was similar using both imputation methods (described below); thus, we report the model results using normal value imputation. Due to a high quantity of missing vital sign data in Cerner (**Supplemental Fig. 1**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>), for clinical model 3, we

limited the cohort to patients with nonmissing vital signs (since missing vital sign data are likely related to the ability of hospitals to electronically capture and report vitals, rather than a decision by clinicians not to measure vitals). Of note, the number of hospitals reporting vital signs increased in later years in the study period, and the characteristics of Adult Sepsis Event patients with and without missing vital signs were generally similar (**Supplemental Table 4**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>).

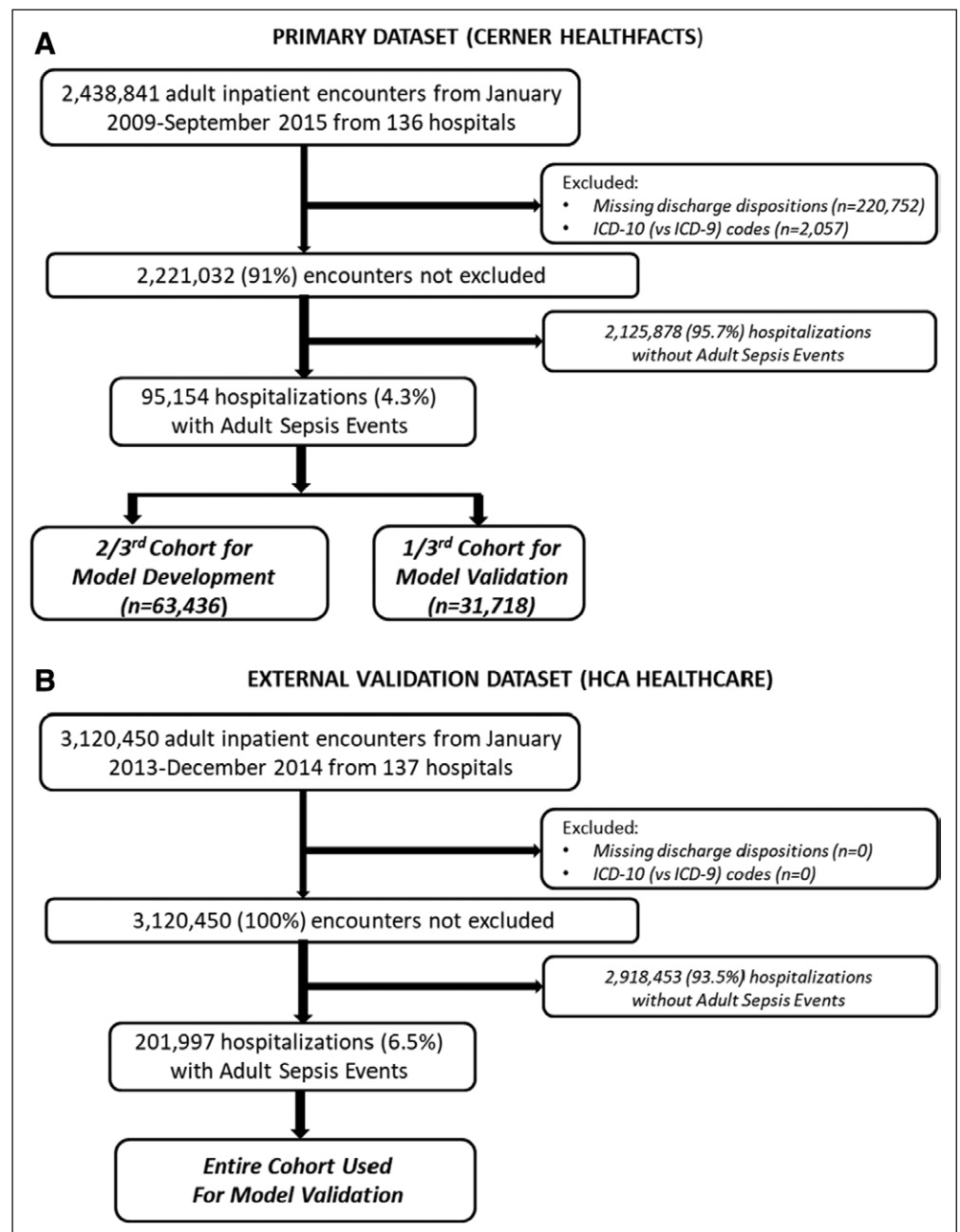
When applying models to the HCA Healthcare dataset, normal values were imputed for missing covariates. We assessed model discrimination using the area under the receiver operating characteristic curve (AUROC) and calibration using the adjusted Brier score (33) and calibration plots. The Hosmer-Lemeshow goodness-of-fit tests were performed on 1,000 random samples of size 1,000 with group size set to 100. Standardized mortality ratios were calculated. Analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC) and R version 3.3.1 (R Core Team, Vienna, Austria; <http://r-project.org>).

## RESULTS

### Study Cohort and Patient Characteristics

Derivation of the study cohort is shown in **Figure 1**. The primary Cerner dataset included 136 hospitals; representation was highest from the South (35%), 65% of

hospitals had less than 200 beds, and 60% were nonteaching hospitals. This cohort included 2,221,032 patients, of whom 95,154 met Adult Sepsis Event criteria and 17,876 died (18.8% mortality). Median age was 68 years (interquartile range, 55–79), and comorbidities were common including diabetes (29%), chronic lung disease (23%), and congestive heart failure (22%) (Table 2). The most common infection sites were urinary tract (27%), lungs (25%), and intra-abdominal space (16%). The most common organ dysfunctions were elevated lactate (40%), acute kidney injury (39%), and hypotension requiring vasopressors (32%). The HCA validation cohort included 201,997 hospitalizations with Adult Sepsis Events.



**Figure 1.** Flowchart for study cohort derivation in primary dataset (Cerner HealthFacts) (A) and external validation dataset (HCA Healthcare) (B). ICD-9 = *International Classification of Diseases*, 9th Edition, ICD-10 = *International Classification of Diseases*, 10th Edition.

**TABLE 2. Characteristics of Sepsis Patients in Primary Cerner HealthFacts Dataset**

Characteristic	All Sepsis Hospitalizations <i>n</i> = 95,154	Discharged Alive <i>n</i> = 77,278	Died In-Hospital <i>n</i> = 17,876
Median age (IQR), yr	68 (55–79)	67 (54–79)	71 (59–81)
Sex, <i>n</i> (%)			
Male (or unknown <sup>a</sup> )	46,288 (48.6)	37,136 (48.1)	9,151 (51.2)
Female	48,867 (51.4)	40,142 (51.9)	8,725 (48.8)
Race, <i>n</i> (%)			
White	69,570 (73.1)	56,854 (73.6)	12,716 (71.1)
Black	18,174 (19.1)	14,464 (18.7)	3,710 (20.8)
Other (or unknown <sup>a</sup> )	7,410 (7.8)	5,960 (7.7)	1,450 (8.1)
Select comorbidities, <sup>b</sup> <i>n</i> (%)			
Cancer	9,973 (10.5)	7,270 (9.4)	2,703 (15.1)
Chronic lung disease	22,282 (23.6)	18,238 (23.6)	4,044 (22.6)
Congestive heart failure	21,285 (22.4)	16,558 (21.4)	4,727 (26.4)
Diabetes	27,994 (29.4)	23,223 (30.1)	4,771 (26.7)
Liver disease	5,664 (6.0)	4,179 (5.4)	1,485 (8.3)
Renal disease	19,202 (20.2)	14,968 (19.4)	4,234 (23.7)
Admission from healthcare facility, <i>n</i> (%)	11,094 (11.7)	8,346 (10.8)	2,748 (15.4)
Select Infectious Syndromes, <sup>c</sup> <i>n</i> (%)			
Pneumonia	24,159 (25.4)	19,110 (24.7)	5,049 (28.2)
Urinary	25,585 (26.9)	22,076 (28.6)	3,509 (19.6)
Intra-abdominal	15,393 (16.2)	12,353 (16.0)	3,040 (17.0)
Skin/soft tissue	7,341 (7.7)	6,526 (8.4)	815 (4.6)
Septicemia/bacteremia	33,954 (35.7)	25,132 (32.5)	8,822 (49.4)
CNS	357 (0.4)	292 (0.4)	65 (0.4)
Obstetric/gynecologic	466 (0.5)	432 (0.6)	34 (0.2)
Hospital-onset sepsis, <i>n</i> (%)	11,534 (12.1)	7,683 (9.9)	3,851 (21.5)
Sepsis organ dysfunction, <sup>d</sup> <i>n</i> (%)			
Vasopressors	30,863 (32.4)	20,273 (26.2)	10,590 (59.2)
Mechanical ventilation	21,182 (22.3)	13,561 (17.8)	7,621 (42.6)
Elevated lactate	38,260 (40.2)	28,268 (36.6)	9,992 (55.9)
Creatinine	36,712 (38.6)	31,747 (41.1)	4,965 (27.8)
Bilirubin	9,267 (9.7)	7,785 (10.1)	1,482 (8.3)
Platelets	11,305 (11.9)	8,724 (11.3)	2,581 (14.4)
Outcomes			
Median hospital LOS (IQR), d	9 (6–15)	9 (6–15)	7 (3–15)
Required ICU admission, <i>n</i> (%)	43,874 (46.1)	32,461 (42.0)	11,413 (63.9)
Median ICU LOS (IQR), d	4 (3–8)	4 (3–7)	4 (2–9)
In-hospital death, <i>n</i> (%)	17,876 (18.8)	0 (0)	17,876 (100)

IQR = interquartile range, LOS = length of stay.

<sup>a</sup>Sex was unknown in 10 patients. Race was unknown in 1,560 patients.

<sup>b</sup>Comorbidities are defined by the Elixhauser index. Cancer includes lymphoma, solid tumor, and metastatic disease.

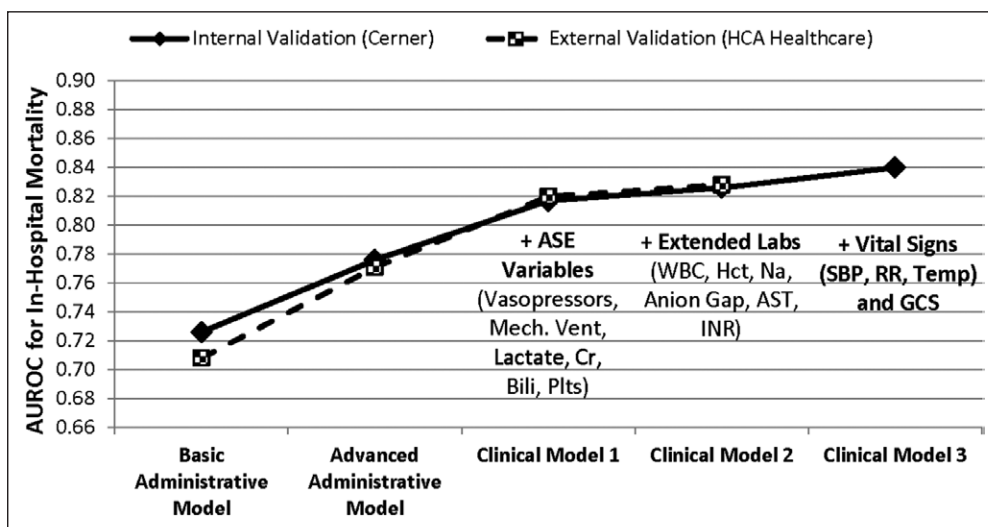
<sup>c</sup>Infectious syndromes were identified using *International Classification of Diseases*, 9th Revision, Clinical Modification discharge diagnosis codes: pneumonia (480.0–480.9, 481, 482.0–482.9, 483.0–483.8, 484.1–484.8, 485, 486), urinary (590.00, 590.01, 590.10, 590.11, 590.2, 590.3, 590.80, 590.81, 590.9, 595.0, 595.2, 595.3, 595.4, 595.89, 595.9, 597.0, 597.80, 597.89, 598.00, 598.01, 599.0), intra-abdominal (008.45, 009.0–009.3, 540.0–540.9, 541, 542, 543.9, 562.01, 562.03, 562.11, 562.13, 567.0–567.9, 569.5, 569.61, 569.71, 569.83, 572.0–572.8, 574.00–574.91, 575.0–575.9, 576.0–576.9, 614.0–614.9), skin/soft tissue (680–686, 035, 376.01, 728.86), septicemia/bacteremia (038.0–038.9, 790.7), CNS (027.0, 036, 320.0–321.1, 321.8, 324.0), and obstetric/gynecologic (614.0–614.5, 616.0–616.1, 616.3–616.4, 634.0, 635.0, 636.0, 637.0, 638.0, 639.0, 646.5, 646.6, 647.9, , 658.4, 659.3, 670, 675).

<sup>d</sup>Sepsis organ dysfunction refers to Adult Sepsis Event organ dysfunction criteria during the sepsis event window (within a ± 2 d window of blood culture day).

## Model Performance

Model performance for discriminating in-hospital mortality is summarized in **Figure 2**. The basic administrative model achieved an AUROC of 0.726 (95% CI, 0.719–0.733) in the Cerner validation cohort. The advanced administrative model achieved an AUROC of 0.776 (95% CI, 0.770–0.783). Adding the clinical data elements required to identify Adult Sepsis Event organ dysfunction further improved performance (clinical model 1; AUROC, 0.817; 95% CI, 0.811–0.822), but there was only modest incremental improvements thereafter with adding more laboratory data (clinical model 2; AUROC, 0.826; 95% CI, 0.820–0.831), and vital signs and GCS (clinical model 3; AUROC, 0.840; 95% CI, 0.833–0.848). Performance of the clinical models was similar when using multiple imputation or normal value imputation for missing laboratory covariates (**Supplemental Table 5**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>). AUROC values in the external HCA Healthcare validation dataset were nearly identical. For example, the AUROC in HCA for clinical model 1 was 0.819 (95% CI, 0.816–0.821) and 0.827 (95% CI, 0.824–0.829 CI) for clinical model 2.

Clinical models 1 and 2 are shown in **Supplemental Tables 6 and 7** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>), whereas the distributions of the clinical variables are shown in **Supplemental Table 8** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>). In general, basic chemistries and blood cell counts were near complete in both datasets (< 5–10% missing), but lactate was missing in 46% of patients in Cerner and 32% in HCA (Supplemental Fig. 1, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>). A variable importance plot illustrating the relative contribution of each set of covariates is shown in **Supplemental Figure 2** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>).



**Figure 2.** Performance of each risk-adjustment model in discriminating for in-hospital mortality among patients with Adult Sepsis Events (ASE). The internal validation cohort was a subset of the Cerner HealthFacts dataset ( $n = 31,718$ ). The external validation cohort was the HCA Healthcare dataset ( $n = 201,997$ ). The results for clinical model 3 shown above are from development and internal validation in Cerner in patients with nonmissing vital signs ( $n = 16,714$  in validation cohort). Vital signs and Glasgow Coma Scale (GCS) were unavailable in the HCA Healthcare validation dataset. AST = aspartate aminotransferase, AUROC = area under the receiver operating curve, Bili = total bilirubin, Cr = creatinine, Hct = hematocrit, INR = international normalized ratio, Mech. Vent = mechanical ventilation, Na = sodium, Plts = platelet count, RR = respiratory rate, SBP = systolic blood pressure, Temp = temperature.

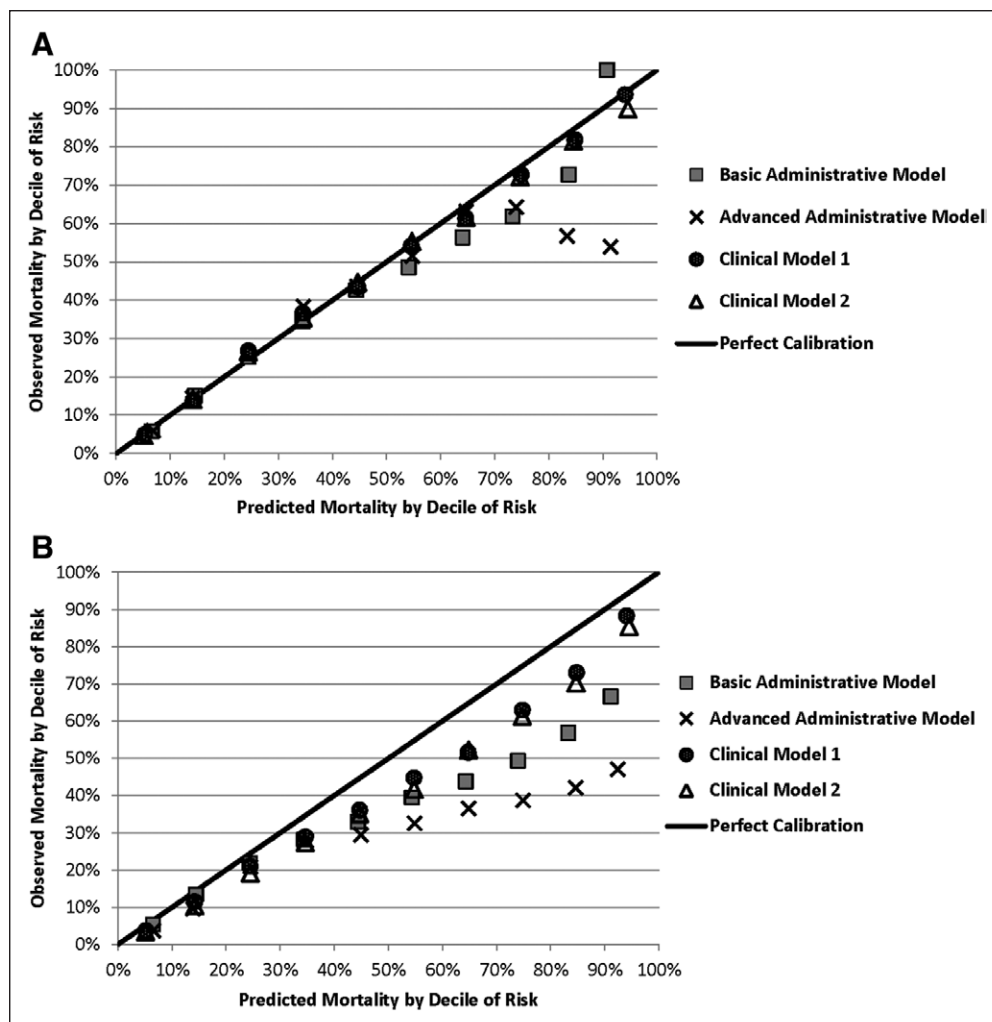
## Model Calibration

Calibration of clinical models 1 and 2 was excellent across all deciles of risk in the Cerner internal validation cohort but less successful in the HCA Healthcare validation cohort, with lower observed mortality rates than predicted (**Fig. 3**). This was in the context of lower overall sepsis mortality in the HCA (15.4%) versus Cerner cohort (18.8%). By comparison, the advanced administrative model performed well at lower risk deciles, but calibration was poor in higher deciles (particularly in the HCA dataset) and worse than the clinical models. Using multiple imputation for missing covariates yielded similar calibration compared with normal value imputation for clinical model 1 in the HCA cohort, but slightly better calibration for clinical model 2 (**Supplemental Fig. 3**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A104>).

## DISCUSSION

As patients, clinicians, administrators, and regulators increasingly focus on the quality of sepsis care provided by hospitals and their sepsis outcomes, hospitals need reliable surveillance methods and robust, generalizable risk-adjustment methods to compare outcomes across time and between institutions. We developed several risk-adjustment models for sepsis mortality based on CDC's Adult Sepsis Event criteria and an array of clinical and administrative data in a large cohort of U.S. hospitals. A model based on the minimum clinical data elements necessary to identify Adult Sepsis Events substantially improved calibration and mortality prediction over administrative models and yielded AUROC values of 0.82, with the successive addition of more complex clinical covariates resulting in diminishing gains. Clinical models performed similarly when validated in an independent hospital system, suggesting good generalizability.

Our study builds on the recent development of CDC's Adult Sepsis Event surveillance definition, which is more sensitive than explicit diagnosis codes for sepsis and has superior positive predictive value compared with "implicit" combinations of infection and organ dysfunction codes (20, 34, 35). Hospital sepsis mortality rates measured using explicit versus implicit administrative definitions have only moderate agreement and yield very different hospital rankings (6). The Adult Sepsis Event definition by contrast minimizes bias from variability in diagnosis, documentation, and coding practices (36). Its simpler criteria also make it easier to automate across diverse EHRs compared with the Sequential Organ Failure Assessment Score and Sepsis-3 criteria, although still retaining good concordance with Sepsis-3 and comparable if not better predictive validity for mortality (9).



**Figure 3.** Calibration plots showing observed versus predicted sepsis in-hospital mortality by decile of risk for various models in internal validation cohort (Cerner HealthFacts) (A) and external validation cohort (HCA Healthcare) (B). Clinical model 1 includes administrative data and basic clinical variables needed to compute Adult Sepsis Event criteria. Clinical model 2 includes the same but adds extended laboratory data. The advanced administrative model includes administrative data and coded severity-of-illness markers.

Other investigators have developed sepsis risk-adjustment models, but each has important limitations for wide-scale implementation for national benchmarking. Although administrative data are readily available, models using clinical data had superior discriminatory performance and calibration at higher deciles of baseline risk compared with the advanced administrative model developed by Ford et al (11). This likely reflects the value of more granular measures of physiology and timing of sepsis onset. Our model may be more generalizable than the New York state sepsis risk-adjustment model because the New York model relies extensively on manually collected data specific to the state's regulatory requirements (14). The Surviving Sepsis Campaign's model was derived in a large, international prospective cohort (13); however, it includes criteria such as "chills with rigor," "bilateral pulmonary infiltrates," and "history suggestive of new infection" that are subjective and inconsistently captured in EHRs.

Our findings provide important insight into the tradeoffs between feasibility of data collection and model performance. Of the various models we created, clinical model 1 is appealing

for operational use because it relies only on the clinical data that hospitals must already gather to identify Adult Sepsis Events. Adding additional laboratory data (clinical model 2) slightly improved performance with a modest increase in complexity. However, calibration across deciles of risk for both clinical models 1 and 2 diminished in the HCA validation dataset versus Cerner cohort, such that observed mortality was generally lower than predicted. This may reflect differences in care patterns and lower sepsis mortality overall in the HCA Healthcare system, which has focused on sepsis quality improvement for the better part of this decade. Adding vital signs and GCS (clinical model 3) provided only small incremental benefit, suggesting that the marginal value of these data may not be commensurate with the effort required to gather and store them.

Indeed, vital signs were frequently missing in Cerner and unavailable in the HCA Healthcare dataset. Conversely, laboratory data such as lactates were more often missing in Cerner. Missing data are clearly an important problem when developing and applying severity-of-illness scores; they can reflect data quality issues, clinical decisions not to test, and/or variability in practice patterns across hospitals. Lactate testing, for example, is likely influenced by provider specialty, unit location at the time of sepsis

onset, and quality improvement initiatives (37, 38). Importantly, we found that missing laboratory values could be handled equally well using multiple imputation versus normal value imputation. This is reassuring since normal value imputation is more feasible for many hospitals. Calibration, however, was slightly worse with clinical model 2 in the HCA dataset using normal value imputation versus multiple imputation. This underscores the challenges inherent to applying risk-adjustment models derived in one dataset to a separate cohort with variable data availability.

Our study has important limitations. We could not perform head-to-head comparisons with the other sepsis severity scores mentioned above or existing ICU severity-of-illness models such as APACHE or SAPS. However, this reflects the practical reality that some of the data needed to calculate these scores are not available as structured data in many hospitals' EHR datasets. More importantly, many septic patients never require ICU care and focusing only on ICU-specific populations risks additional confounding from variability in hospitals' ICU capacities and admission thresholds (18, 19). We did not have data on mortality

occurring after hospital discharge. Our datasets did not include insurance status, which may be an important predictor of sepsis mortality (11, 12, 14). Our models relied on comorbidities and infectious diagnoses, which may be variably diagnosed and coded across institutions. We only compared model performance using EHR versus administrative data and did not assess the potential impact of different risk-adjustment methods on hospitals' sepsis mortality rankings. This is an important area for future research. Missing data were common in both datasets, with variable quantities of missing laboratory data and unavailable vital signs in the HCA Healthcare dataset. We used ICD-9-CM codes in our models; further research is needed to refine and update our models to use ICD-10 data. Real-world implementation of our risk-adjustment models for hospital benchmarking would require periodic updating and recalibration over time, ideally using nationally representative datasets with minimal missing data. This would be analogous to the periodic updates applied to risk-adjustment models and baseline standardized infection ratio levels for hospitals' healthcare-associated infection rates reported to CDC's National Healthcare Safety Network (39).

Finally, Adult Sepsis Event criteria do not perfectly match Sepsis-3 criteria; one prior study demonstrated 70% sensitivity based on expert clinician-adjudicated medical record reviews (20). Of note, however, the Sepsis-3 cases missed by Adult Sepsis Events tend to be mild infections with low risk of mortality (9, 20). Furthermore, there is no true "gold standard" for sepsis and there is increasing recognition that clinical care and surveillance merit different definitions (40). Adult Sepsis Events were created and optimized for widespread surveillance rather than clinical care and as such are easier to apply to routine EHR data than Sepsis-3 and the Sequential Organ Failure Assessment score (9).

In conclusion, models based on administrative and EHR data accurately predict hospital mortality for patients with Adult Sepsis Events and outperform models based on administrative data alone. A risk-adjustment model incorporating laboratory test results, vasopressors, and mechanical ventilation without vital signs may achieve a good balance between data collection needs and performance, but EHR-based models must be attentive to potential variability in data quality and availability. With ongoing testing and refinement of these risk-adjustment models, Adult Sepsis Event surveillance may enable more meaningful comparisons of hospital sepsis outcomes and provide an important new window into quality of care.

## REFERENCES

- Venkatesh AK, Slesinger T, Whittle J, et al: Preliminary performance on the new CMS sepsis-1 national quality measure: Early insights from the Emergency Quality Network (E-QUAL). *Ann Emerg Med* 2018; 71:10–15. e11
- Barbash IJ, Davis B, Kahn JM: National performance on the Medicare SEP-1 sepsis quality measure. *Crit Care Med* 2019; 47:1026–1032
- Klompas M, Rhee C: The CMS sepsis mandate: Right disease, wrong measure. *Ann Intern Med* 2016; 165:517–518
- Wang HE, Donnelly JP, Shapiro NI, et al: Hospital variations in severe sepsis mortality. *Am J Med Qual* 2015; 30:328–336
- Prescott HC, Kepreos KM, Wiitala WL, et al: Temporal changes in the influence of hospitals and regional healthcare networks on severe sepsis mortality. *Crit Care Med* 2015; 43:1368–1374
- Walkey AJ, Shieh MS, Liu VX, et al: Mortality measures to profile hospital performance for patients with septic shock. *Crit Care Med* 2018; 46:1247–1254
- Rhee C, Jentzsch MS, Kadri SS, et al; Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program: Variation in identifying sepsis and organ dysfunction using administrative versus electronic clinical data and impact on hospital outcome comparisons. *Crit Care Med* 2019; 47:493–500
- Centers for Disease Control and Prevention: Hospital Toolkit for Adult Sepsis Surveillance. 2018. Available at: [https://www.cdc.gov/sepsis/pdfs/Sepsis-Surveillance-Toolkit-Mar-2018\\_508.pdf](https://www.cdc.gov/sepsis/pdfs/Sepsis-Surveillance-Toolkit-Mar-2018_508.pdf). Accessed August 1, 2019
- Rhee C, Zhang Z, Kadri SS, et al; CDC Prevention Epicenters Program: Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 sequential organ failure assessment criteria. *Crit Care Med* 2019; 47:307–314
- Rhee C, Dantes RB, Epstein L, et al: Using objective clinical data to track progress on preventing and treating sepsis: CDC'S new 'adult sepsis event' surveillance strategy. *BMJ Qual Saf* 2019; 28:305–309
- Ford DW, Goodwin AJ, Simpson AN, et al: A severe sepsis mortality prediction model and score for use with administrative data. *Crit Care Med* 2016; 44:319–327
- Lagu T, Lindenauer PK, Rothberg MB, et al: Development and validation of a model that uses enhanced administrative data to predict mortality in patients with sepsis. *Crit Care Med* 2011; 39:2425–2430
- Osborn TM, Phillips G, Lemeshow S, et al: Sepsis severity score: An internationally derived scoring system from the surviving sepsis campaign database. *Crit Care Med* 2014; 42:1969–1976
- Phillips GS, Osborn TM, Terry KM, et al: The New York sepsis severity score: Development of a risk-adjusted severity model for sepsis. *Crit Care Med* 2018; 46:674–683
- Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13:818–829
- Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957–2963
- Lemeshow S, Teres D, Klar J, et al: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270:2478–2486
- Rohde JM, Odden AJ, Bonham C, et al: The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit. *J Hosp Med* 2013; 8:243–247
- Wunsch H, Rowan KM, Angus DC: International comparisons in critical care: A necessity and challenge. *Curr Opin Crit Care* 2007; 13:725–731
- Rhee C, Dantes R, Epstein L, et al: Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA* 2017; 318:1241–1249
- Choudhry SA, Li J, Davis D, et al: A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model. *Online J Public Health Inform* 2013; 5:219
- Goyal A, Spertus JA, Gosch K, et al: Serum potassium levels and mortality in acute myocardial infarction. *JAMA* 2012; 307:157–164
- Lagu T, Pekow PS, Shieh MS et al: Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure. *Circ Heart Fail* 2016; 9:e002912
- Petrick JL, Nguyen T, Cook MB: Temporal trends of esophageal disorders by age in the Cerner health facts database. *Ann Epidemiol* 2016; 26:151–154. e4
- Septimus E, Hickok J, Moody J, et al: Closing the translation gap: Toolkit-based implementation of universal decolonization in adult intensive care units reduces central line-associated bloodstream infections in 95 community hospitals. *Clin Infect Dis* 2016; 63:172–177
- Elixhauser A, Steiner C, Harris DR, et al: Comorbidity measures for use with administrative data. *Med Care* 1998; 36:8–27
- Christensen KL, Holman RC, Steiner CA, et al: Infectious disease hospitalizations in the United States. *Clin Infect Dis* 2009; 49:1025–1035
- Rhee C, Gohil S, Klompas M: Regulatory mandates for sepsis care—reasons for caution. *N Engl J Med* 2014; 370:1673–1676



29. Liang K, Zeger SL: Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73:13–22
30. Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; 315:762–774
31. Vincent JL, de Mendonça A, Cantraine F, et al: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Working group on “sepsis-related problems” of the European Society of Intensive Care Medicine. *Crit Care Med* 1998; 26:1793–1800
32. Seymour CW, Kahn JM, Cooke CR, et al: Prediction of critical illness during out-of-hospital emergency care. *JAMA* 2010; 304:747–754
33. 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
34. Whittaker SA, Mikkelsen ME, Galeski DF, et al: Severe sepsis cohorts derived from claims-based strategies appear to be biased toward a more severely ill patient population. *Crit Care Med* 2013; 41:945–953
35. Jolley RJ, Sawka KJ, Yergens DW, et al: Validity of administrative data in recording sepsis: A systematic review. *Crit Care* 2015; 19:139
36. Simpson SQ: Surveillance for adult sepsis events: An idea whose time has come. *Crit Care Med* 2019; 47:467–468
37. Rhee C, Murphy MV, Li L, et al; Centers for Disease Control and Prevention Prevention Epicenters Program: Lactate testing in suspected sepsis: Trends and predictors of failure to measure levels. *Crit Care Med* 2015; 43:1669–1676
38. Whippy A, Skeath M, Crawford B, et al: Kaiser permanente's performance improvement system, part 3: Multisite improvements in care for patients with sepsis. *Jt Comm J Qual Patient Saf* 2011; 37:483–493
39. Centers for Disease Control and Prevention: The NHSN Standardized Infection Ratio (SIR). Updated March 2019. Available at: <https://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/nhsn-sir-guide.pdf>. Accessed August 1, 2019
40. Angus DC, Seymour CW, Coopersmith CM, et al: A framework for the development and interpretation of different sepsis definitions and clinical criteria. *Crit Care Med* 2016; 44:e113–e121