

# Impact of Risk Adjustment Using Clinical vs Administrative Data on Hospital Sepsis Mortality Comparisons

Chanu Rhee,<sup>1,2</sup> Zhonghe Li,<sup>1</sup> Rui Wang,<sup>1</sup> Yue Song,<sup>1,3</sup> Sameer S. Kadri,<sup>4</sup> Edward J. Septimus,<sup>1,5</sup> Huai-Chun Chen,<sup>6</sup> David Fram,<sup>6</sup> Robert Jin,<sup>1</sup> Russell Poland,<sup>1,7</sup> Kenneth Sands,<sup>1,7</sup> and Michael Klompas<sup>1,2</sup>; for the CDC Prevention Epicenters Program

<sup>1</sup>Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA, <sup>2</sup>Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA, <sup>3</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA, <sup>4</sup>Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, Massachusetts, USA, <sup>5</sup>Texas A&M Health Science Center College of Medicine, Houston, Texas, USA, <sup>6</sup>Commonwealth Informatics, Waltham, Massachusetts, USA, and <sup>7</sup>Clinical Services Group, HCA Healthcare, Nashville, Tennessee, USA

**Background.** A reliable risk-adjusted sepsis outcome measure could complement current national process metrics by identifying outlier hospitals and catalyzing additional improvements in care. However, it is unclear whether integrating clinical data into risk adjustment models identifies similar high- and low-performing hospitals compared with administrative data alone, which are simpler to acquire and analyze.

**Methods.** We ranked 200 US hospitals by their Centers for Disease Control and Prevention Adult Sepsis Event (ASE) mortality rates and assessed how rankings changed after applying (1) an administrative risk adjustment model incorporating demographics, comorbidities, and codes for severe illness and (2) an integrated clinical and administrative model replacing severity-of-illness codes with laboratory results, vasopressors, and mechanical ventilation. We assessed agreement between hospitals' risk-adjusted ASE mortality rates when ranked into quartiles using weighted kappa statistics ( $\kappa$ ).

**Results.** The cohort included 4 009 631 hospitalizations, of which 245 808 met ASE criteria. Risk-adjustment had a large effect on rankings: 22/50 hospitals (44%) in the worst quartile using crude mortality rates shifted into better quartiles after administrative risk adjustment, and a further 21/50 (42%) of hospitals in the worst quartile using administrative risk adjustment shifted to better quartiles after incorporating clinical data. Conversely, 14/50 (28%) hospitals in the best quartile using administrative risk adjustment shifted to worse quartiles with clinical data. Overall agreement between hospital quartile rankings when risk-adjusted using administrative vs clinical data was moderate ( $\kappa = 0.55$ ).

**Conclusions.** Incorporating clinical data into risk adjustment substantially changes rankings of hospitals' sepsis mortality rates compared with using administrative data alone. Comprehensive risk adjustment using both administrative and clinical data is necessary before comparing hospitals by sepsis mortality rates.

**Keywords.** Adult Sepsis Event; hospital comparisons; outcome measure; risk adjustment; sepsis.

Sepsis is a leading cause of death, disability, and cost to the health care system [1, 2]. The high burden of sepsis has spurred national efforts to improve timely treatment and bundle adherence, including the Centers for Medicare & Medicaid Services (CMS) Severe Sepsis/Septic Shock Early Management Bundle (SEP-1) that was implemented in 2015 [3, 4]. CMS is now expanding beyond process metrics and developing a sepsis

outcome measure for potential use in future quality and payment programs [5]. A reliable sepsis outcome measure could catalyze additional improvements in sepsis care by identifying best practices, flagging poor performers, and motivating continuous improvement [6].

Prior work has demonstrated substantial variability in diagnosis and coding for sepsis and organ dysfunction among clinicians and hospitals, suggesting that administrative data are not suitable for anchoring comparisons of hospitals' sepsis outcomes [7, 8]. In 2018, the US Centers for Disease Control and Prevention (CDC) released the "Adult Sepsis Event" (ASE) surveillance definition that uses clinical data routinely available in electronic health record systems (EHRs) to identify sepsis [9]. The ASE is modeled after the International Consensus Sepsis-3 definition but is optimized for automated surveillance across a wide array of hospitals and settings by using streamlined infection and organ dysfunction criteria that can be easily applied using data commonly found in most hospitals' EHRs [10, 11].

Received 14 April 2020; editorial decision 28 May 2020; accepted 1 June 2020.

Correspondence: C. Rhee, MD, MPH, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401, Boston, MA 02215 (crhee@bwh.harvard.edu).

Open Forum Infectious Diseases®

© The Author(s) 2020. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com  
DOI: 10.1093/ofid/ofaa213

The ASE therefore provides an objective and consistent way to identify sepsis cases across hospitals. However, robust risk adjustment is still required to make fair comparisons of hospitals' sepsis mortality rates as sepsis mortality is influenced by many factors beyond the care rendered, including patients' demographics, comorbidities, severity of illness, and site of infection [12]. As an example, an elderly patient with end-stage cancer presenting with septic shock from a perforated bowel has a substantially higher risk of mortality than a young woman presenting with hypotension and acute kidney injury from pyelonephritis, regardless of quality of care.

We recently developed risk adjustment models for ASE that utilize EHR clinical data and showed that these have better discrimination for mortality and calibration than models based on administrative data alone [13]. It is unclear, however, whether integrating clinical data into risk adjustment models adds substantial marginal value for hospital comparisons over adjustments using administrative data alone, which are simpler to acquire and analyze. If risk adjustment using administrative models identifies similar high- or low-performing hospitals compared with EHR models, then there is little basis to justify the added complexity of integrating clinical data. On the other hand, if integrating clinical data with administrative data generates substantially different rankings, this would suggest that the extra complexity is necessary for credible benchmarking.

In this study, we assessed the impact of risk adjustment on relative rankings of hospitals' ASE mortality rates when applying administrative vs administrative plus clinical risk adjustment models. We further compared risk-adjusted hospital rankings when sepsis was defined using administrative codes vs ASE clinical criteria.

## METHODS

### Study Design, Data Sources, and Sepsis Case Definition

We conducted a retrospective study of adults aged  $\geq 20$  years who were admitted in calendar years 2013 or 2014 to 200 US acute care hospitals drawn from 3 data sets: Cerner HealthFacts, HCA Healthcare, and the Institute of Health Metrics. These data sets were previously used in a national epidemiologic study of sepsis and include a geographically diverse mix of academic and community hospitals [1].

We identified sepsis using the CDC's ASE definition, which identifies hospitalizations with presumed serious infection (blood culture order and new antibiotics continued for  $\geq 4$  days or until  $\leq 1$  day before 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) [9].

We excluded hospitals with  $< 50$  ASE cases during the study period due to the uncertainty associated with rates estimated

from low-volume hospitals, similar to prior studies comparing hospital sepsis outcomes [8, 14]. We further excluded encounters with missing discharge dispositions and those with *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM) instead of ICD-9-CM codes [13].

### Risk Adjustment Methods and Analyses

Our outcome of interest was all-cause in-hospital mortality. In our primary analysis, we applied 2 sets of risk adjustment methods to patients identified by ASE criteria (Table 1). First, we applied an administrative model developed by Ford et al. that incorporates demographics, comorbidities, and ICD-9-CM codes indicative of severity of illness on admission (mechanical ventilation, shock, hemodialysis, and intensive care unit [ICU] admission) [15]. In a previous analysis, this administrative model achieved an area under the receiver operating characteristics curve (AUROC) of 0.776 (95% CI, 0.770–0.783) in the Cerner data set and 0.771 (95% CI, 0.768–0.773) in the HCA Healthcare data set, with good calibration at low deciles of baseline risk but worse calibration with higher-risk patients [13]. Second, we applied a clinical model that uses similar administrative data (demographics and comorbidities) but also adds site of infection (by ICD-9-CM codes) and days in the hospital until sepsis onset and uses physiologic data rather than codes for severity of illness (laboratory data from chemistries, complete blood cell counts, and liver function tests; vasopressors; and mechanical ventilation). This model had an AUROC of 0.826 (95% CI, 0.820–0.831) in Cerner and 0.827 (95% CI, 0.824–0.829) in the HCA

**Table 1. Components of Adult Sepsis Event Risk Adjustment Models: Administrative Model vs Integrated Administrative and Clinical Model**

Model Components	Administrative Model <sup>a</sup>	Integrated Administrative and Clinical Model <sup>b</sup>
<b>Administrative data</b>		
Demographics	✓	✓
Comorbidities	✓	✓
Mechanical ventilation	✓	✓
Shock code	✓	
Hemodialysis code	✓	
ICU admission	✓	✓
Infection site		✓
<b>Clinical data</b>		
Vasopressors		✓
Laboratory data: chemistries		✓
Complete blood cell counts, liver function tests, lactate		
Days to sepsis onset		✓

Abbreviations: ICU, intensive care unit.

<sup>a</sup>Administrative data were based on encounter data and ICD-9-CM codes.

<sup>b</sup>In the integrated administrative and clinical model, mechanical ventilation, ICU admission, vasopressors, and laboratory data within  $\pm 1$  calendar day of the day of sepsis onset were used. The day of sepsis onset was defined as the earliest day the blood culture or first qualifying antibiotic day occurred. The worst values for laboratory values within that window were used. Missing values were assumed to be normal.

Healthcare data set, with good calibration across all baseline risk deciles [13]. Missing values were imputed using normal values, as we had previously shown that normal value imputation and multiple imputation generate similar model performance [13].

We applied risk adjustment using both methods to calculate each hospital's expected number of deaths, then divided the hospital-level observed number of deaths by the expected number of deaths to obtain risk-adjusted standardized mortality ratios for each hospital [16]. Using this method, lower ratios indicate better risk-adjusted outcomes than higher ratios.

We ranked all hospitals by sepsis-standardized mortality ratios and calculated the number of hospitals in the worst quartile using crude mortality rates that shifted to better quartiles using administrative risk adjustment, and then the number of administratively adjusted hospitals in the worst quartile that shifted to better quartiles using clinical risk adjustment. We further assessed the agreement between hospitals' crude, administrative-adjusted, and clinical-adjusted relative rankings, divided into quartiles within the study cohort, using a weighted kappa statistic ( $\kappa$ ). A weighted  $\kappa$  is used to calculate agreement for ordinal ratings and gives "partial" credit for close but imperfect agreement [17]. The correlation between hospitals' standardized mortality ratios using different risk adjustment methods was also assessed using the Spearman correlation coefficient ( $\rho$ ). As per common convention, we a priori classified agreement and correlation as very strong for  $\kappa$  and  $\rho$  values  $\geq 0.90$ , strong from 0.70 to 0.89, moderate from 0.40 to 0.69, weak from 0.10 to 0.39, and negligible from 0.00 to 0.10 [18]. We also explored whether several basic hospital characteristics, including size (<200 beds, 200–499 beds,  $\geq 500$  beds), region (Northeast, Midwest, South, West), and teaching status, were associated with a change in quartile ranking based on clinical vs administrative risk adjustment using a logistic regression model [19].

In addition to comparing correlation and agreement between clinical-adjusted vs administrative-adjusted ASE mortality, we compared hospital-standardized mortality ratios using administrative sepsis definitions after risk adjustment using the administrative model described above. We reasoned that if using administrative data for both sepsis identification and risk adjustment yields similar results compared with using ASE criteria with clinical risk adjustment, this would argue against the need to use clinical data at all for hospital comparisons. We used 2 administrative definitions for these comparisons: 1) explicit severe sepsis (ICD-9-CM code 995.92) or septic shock (785.52) codes and 2) either implicit codes for infection and organ dysfunction or explicit severe sepsis/septic shock codes alone (modified Angus criteria) [20, 21]. Compared with implicit sepsis codes, explicit sepsis codes tend to have lower sensitivity and higher positive predictive values and identify a more severely ill cohort of patients [1, 21, 22].

Analyses were conducted using SAS, version 9.4 (SAS Institute, Cary, NC, USA). The study was approved with a

waiver of informed consent by the Institutional Review Board at Harvard Pilgrim Health Care Institute.

## RESULTS

### Study Cohort Hospitals and Case Counts

The study cohort included 200 hospitals. Most hospitals were medium-sized ( $n = 105$ , 52.5%), nonteaching ( $n = 128$ , 64%), and from the South ( $n = 120$ , 60%) (Table 2). There were 4 009 631 adult patients admitted to these hospitals during 2013–2014, including 245 808 sepsis hospitalizations by ASE criteria (median [interquartile range {IQR}], 977.5 [489–1786] per hospital). The crude mortality rate of patients meeting ASE criteria was 15.5% and ranged from 5.4% to 34.8% across hospitals (median [IQR], 15.4% [12.7%–17.8%]).

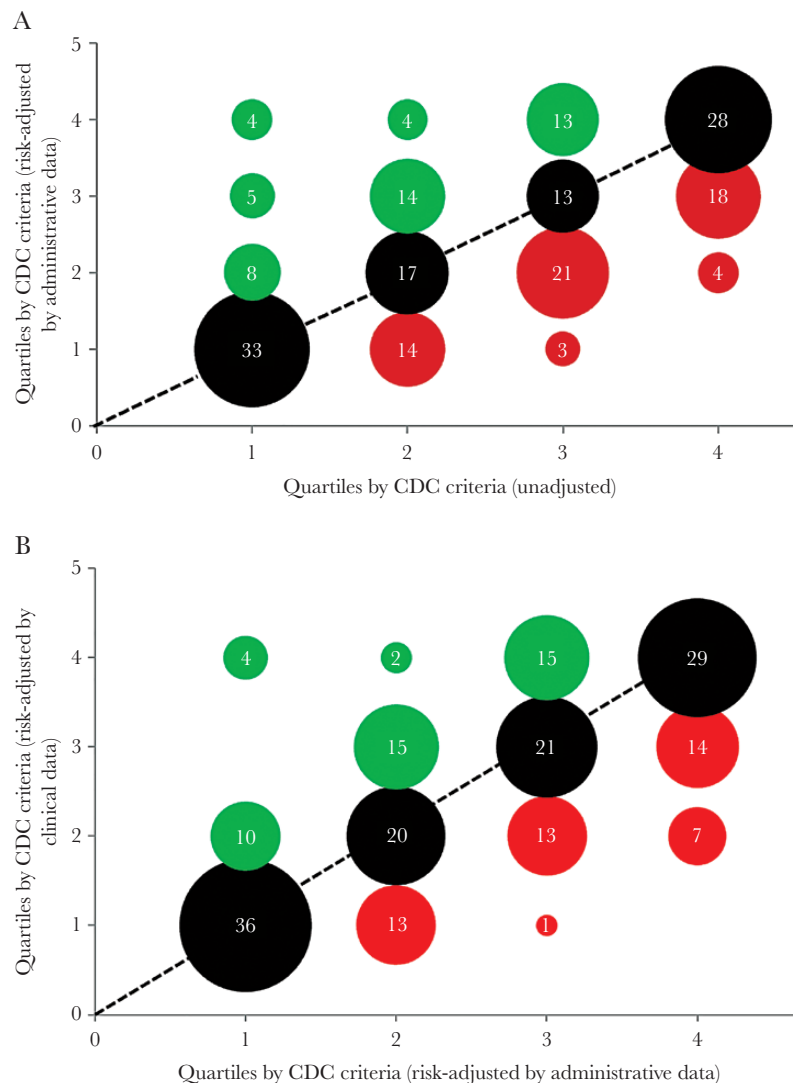
### Impact of Risk Adjustment Using Administrative Data vs Integrated Clinical Data

Risk adjustment had a large effect on rankings: 22/50 hospitals (44%) in the worst quartile using crude mortality rates shifted into better quartiles after administrative risk adjustment (Figure 1A). A further 21/50 (42%) hospitals in the worst quartile using administrative risk adjustment shifted to better quartiles after incorporating clinical data (Figure 1B). Conversely, 17/50 (34%) hospitals in the best quartile using crude mortality rates shifted to lower quartiles after adjusting for administrative data, and a further 14/50 (28%) hospitals in the best quartile per administrative risk adjustment shifted to worse quartiles after integrating clinical data. The correlation between hospital standardized mortality ratios when risk-adjusting using clinical vs administrative data was strong ( $\rho$ , 0.72; 95% CI, 0.65–0.78) (Figure 2A), but the overall agreement between hospital quartile rankings was only moderate ( $\kappa$ , 0.55; 95% CI, 0.47–0.63).

**Table 2. Characteristics of Study Hospitals**

Characteristic	No. (%) or Median (IQR)
<b>Hospital size</b>	
Small (<200 beds)	78 (39.0)
Medium (200–499 beds)	105 (52.5)
Large ( $\geq 500$ beds)	17 (8.5)
<b>Region</b>	
Northeast	19 (9.5)
Midwest	18 (9.0)
South	120 (60.0)
West	43 (21.5)
<b>Teaching status</b>	
Teaching	72 (36.0)
Nonteaching	128 (64.0)
<b>Case counts, 2013–2014</b>	
Hospitalizations	17 197.5 (9208.5–28 528.5)
Sepsis cases by Adult Sepsis Event criteria	977.5 (489–1786)
Explicit sepsis cases	472.5 (228–735)
Implicit sepsis cases	1924.5 (998–3304)

Abbreviation: IQR, interquartile range.



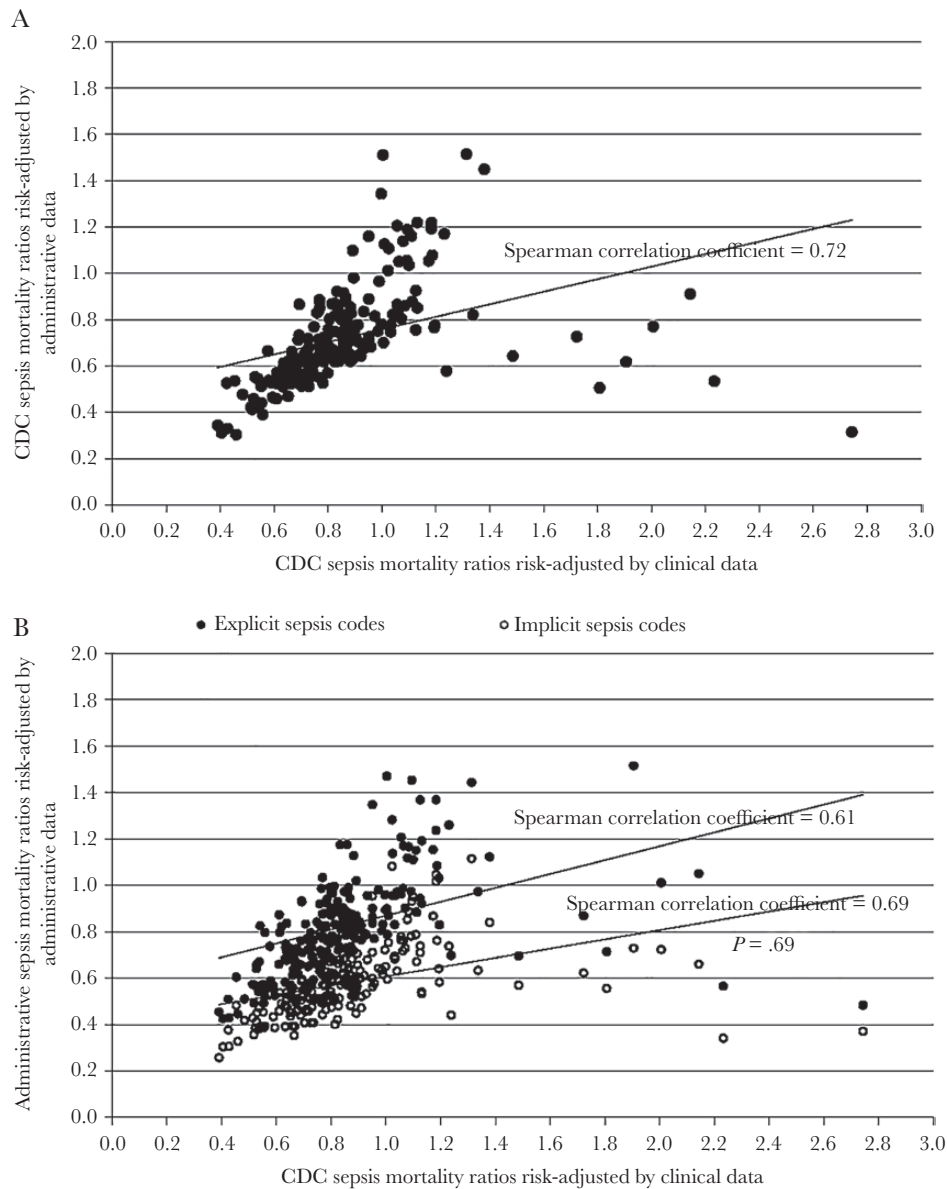
**Figure 1.** Concordance of hospital Centers for Disease Control and Prevention Adult Sepsis Event sepsis mortality rates when ranked into quartiles: (A) unadjusted vs risk-adjusted by administrative data, (B) risk-adjusted using administrative data vs integrated clinical data. The figure shows the impact of risk adjustment on hospitals' observed Adult Sepsis Event mortality rankings. Bubble sizes are proportional to the number of hospitals in each matched quartile. The actual number of hospitals in each category is denoted within the bubbles. The cohort included 200 hospitals. Lower quartiles indicate better performance (ie, quartile 1 = lowest sepsis mortality rates, quartile 4 = highest mortality rates). The bubbles in black, connected by the dotted lines, indicate where all hospitals would lie if concordance was perfect between the various comparisons. Red bubbles below the dotted line indicate cases in which hospitals' unadjusted sepsis mortality rankings shift into better quartiles after risk adjustment by the administrative model (Figure 1A), or in which hospitals' administrative risk-adjusted mortality rankings shift into better quartiles after risk-adjustment by the integrated administrative and clinical model (Figure 1B). Bubbles in green above the dotted line indicate the opposite. For example, Figure 1A shows that 22 (18 + 4) hospitals that were ranked in the worst quartile of unadjusted sepsis mortality rates shifted to better quartiles after risk adjustment using the administrative model. Figure 1B shows that 21 (14 + 7) hospitals in the worst quartile of sepsis mortality after risk adjustment by the administrative model shifted to better quartiles after risk adjustment by the integrated administrative and clinical model.

We did not detect a significant association between a change in quartile rankings based on clinical vs administrative risk adjustment in any of the hospital characteristics we examined.

#### Comparison of Hospital Sepsis Mortality Rates by Adult Sepsis Events vs Administrative Sepsis Definitions

The correlation of unadjusted sepsis mortality rates using ASE criteria was moderate compared with explicit administrative definitions ( $\rho$ , 0.61; 95% CI, 0.51–0.69) and compared with

implicit sepsis codes ( $\rho$ , 0.69; 95% CI, 0.61–0.76). The correlation was similar after risk adjustment using clinical data for ASE criteria and administrative data for administrative definitions (explicit sepsis codes:  $\rho$ , 0.68; 95% CI, 0.59–0.75; implicit codes:  $\rho$ , 0.70; 95% CI, 0.62–0.76) (Figure 2B). Agreement between unadjusted hospital mortality quartile rankings by ASE criteria and administrative definitions was moderate (explicit sepsis codes:  $\kappa$ , 0.40; 95% CI, 0.31–0.49; implicit codes:  $\kappa$ , 0.50; 95% CI, 0.41–0.58) and remained moderate



**Figure 2.** Correlation between hospital standardized sepsis mortality ratios: (A) Centers for Disease Control and Prevention (CDC) Adult Sepsis Event risk-adjusted using clinical vs administrative data, (B) clinical-adjusted CDC Adult Sepsis Event mortality ratios vs administrative-adjusted sepsis diagnosis codes. “Explicit” sepsis codes include *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes for severe sepsis (995.92) or septic shock (785.52). “Implicit” sepsis codes include combinations of ICD-9-CM codes for infection and organ dysfunction or explicit sepsis codes alone.

after risk adjustment using clinical data for ASE and administrative data for administrative definitions (explicit codes:  $\kappa$ , 0.52; 95% CI, 0.43–0.61; implicit codes:  $\kappa$ , 0.54; 95% CI, 0.46–0.63). Eighteen of 50 (36%) hospitals in the worst quartile by administrative-adjusted explicit sepsis codes shifted to better quartiles by clinical-adjusted ASE criteria, while 17/50 (34%) hospitals in the best quartile shifted to worse quartiles. Similarly, 19/50 (38%) hospitals in the worst quartile by administrative-adjusted implicit sepsis codes shifted to better quartiles by clinical-adjusted ASE criteria, while 19/50 (38%) hospitals in the best quartile shifted to worse quartiles.

## DISCUSSION

In this study, we demonstrated that risk adjustment has a large impact on hospitals’ rankings for sepsis mortality rates, as 44% of hospitals in the worst quartile using crude data shifted to higher quartiles after administrative risk adjustment. Moreover, adding clinical data to administrative models and replacing severity-of-illness codes with physiologic data had a further large impact on hospitals’ sepsis-mortality rankings: 42% of hospitals deemed to be in the worst quartile of performance by administrative risk adjustment shifted to more moderate quartiles by clinical risk adjustment. Finally, hospital rankings based



entirely on administrative data for sepsis identification and risk adjustment had only moderate agreement with rankings by Adult Sepsis Event criteria after clinical risk adjustment.

Sepsis care and outcomes are subjects of intense interest for regulators, payors, quality advocates, and hospitals. CMS and many states now require hospitals to publicly report compliance with sepsis management bundles. Ultimately, however, the goal of sepsis management bundles is to improve sepsis outcomes, and thus there are efforts currently underway by CMS to develop an outcome measure that can be used to further compare and contrast hospitals [5]. Prior studies have demonstrated substantial variability across hospitals in the accuracy of administrative definitions of sepsis, underscoring the importance of anchoring hospital sepsis comparisons to consistent clinical criteria such as the CDC's Adult Sepsis Event definition [8, 23]. These analyses, however, left unanswered the question of how to credibly risk-adjust hospitals' sepsis-associated mortality rates.

The large effect of risk adjustment using administrative data on hospitals' sepsis mortality rankings is unsurprising given that there is considerable variation between hospitals in the complexity and level of illness of patients they manage [24]. However, our finding that risk adjustment using clinical vs administrative data alone leads to substantial further changes in hospital rankings has important implications for the operationalization of a sepsis outcome measure. These findings are concordant with a recent analysis that demonstrated major discrepancies in mortality trends for patients hospitalized with heart failure and pneumonia when risk-adjusting using administrative vs clinical risk adjustment variables [25]. Differences in hospital rankings for surgical site infection rates have also been demonstrated when risk-adjusting using clinical vs administrative data [26].

Our findings contrast with a study by Darby et al. that developed an administrative risk adjustment model for sepsis using data from Pennsylvania and showed that correlation with risk-adjusted mortality rates was very high when adding clinical data, with minimal change in observed hospital performance [27]. Their approach differs from ours, however, as the Darby analysis used implicit sepsis codes rather than clinical criteria to identify sepsis and added laboratory values on admission to an administrative model that already captured physiologic severity of illness through organ failure codes. In contrast, our clinical model replaced all physiologic severity-of-illness codes in the administrative model with an array of clinical data, including laboratory results, days to sepsis onset, vasopressor administration, and need for mechanical ventilation on the day of sepsis onset. This is an important distinction, as prior work has demonstrated substantial variability in hospitals' threshold for diagnosing and coding for organ dysfunction [7].

The practical impact on hospital profiling that we observed when risk-adjusting CDC clinical criteria with the administrative vs clinical model is perhaps surprising as the administrative

risk adjustment model performed nearly as well in these data sets for discriminating mortality, with AUROC values of ~0.77 compared with 0.82 for the clinical model [13]. This suggests that model discrimination is not the only important factor when considering the utility of risk adjustment models; researchers and policy makers need to take into account the concordance of simpler vs more sophisticated models in identifying outliers. Interestingly, none of the basic hospital characteristics that were examined in our data set predicted changes in rankings when using clinical risk adjustment.

Our study has important limitations. First, we used a convenience sample of hospitals, which may limit the generalizability of our findings. However, our cohort included a large number of geographically diverse academic and community hospitals of varying sizes. Second, we did not perform direct comparisons with other existing ICU severity-of-illness models that use clinical data, such as APACHE or SAPS, as some of the data needed to calculate these scores are not available in many hospitals' EHR data sets. Furthermore, many patients with sepsis are treated outside of the ICU, and focusing only on ICU populations could create bias due to variability in hospitals' ICU bed capacities and thresholds to admit to the ICU [28, 29]. Third, we did not have data on mortality that might occur shortly after hospital discharge and so focused only on in-hospital death as an outcome. Fourth, based on data availability, we only examined basic hospital characteristics and their associations with changes in hospital rankings when using clinical vs administrative risk adjustment, and we cannot rule out the importance of other factors such as hospital ownership, specific diagnosis, and coding practices within health care systems, safety net status, and more. Lastly, the applicability of our findings in the ICD-10 era is unknown and warrants additional research. It is unlikely, however, that the shift from ICD-9 to ICD-10 decreased variation in coding given ICD-10's greater complexity and use of multidirectional mappings [30, 31].

In conclusion, risk adjustment for sepsis mortality substantially changes hospitals' relative rankings, and integrating clinical data into risk adjustment models generates very different rankings compared with using administrative data alone. Comprehensive risk adjustment using both administrative and clinical data is necessary before comparing hospitals' sepsis mortality rates.

## Acknowledgments

**Financial support.** This work was supported by the Centers for Disease Control and Prevention (grant number U54CK000484), the Agency for Healthcare Research and Quality (K08HS025008 to C.R.), and intramural funds from the National Institutes of Health Clinical Center and National Institute of Allergy and Infectious Diseases (S.S.K.).

**Disclaimer.** The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the National Institutes of Health.

**Potential conflicts of interest.** None of the authors has any conflicts of interest to declare. All authors: no reported conflicts of interest. All authors

have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. Rhee C, Dantes R, Epstein L, et al; CDC Prevention Epicenter Program. Incidence and trends of sepsis in us hospitals using clinical vs claims data, 2009-2014. *JAMA* **2017**; 318:1241-9.
2. Liu V, Escobar GJ, Greene JD, et al. Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA* **2014**; 312:90-2.
3. Barbash IJ, Davis B, Kahn JM. National performance on the Medicare SEP-1 sepsis quality measure. *Crit Care Med* **2018**.
4. Rhee C, Chiotos K, Cosgrove SE, Het al. Infectious Diseases Society of America position paper: recommended revisions to the national severe sepsis and septic shock early management bundle (SEP-1) sepsis quality measure. *Clin Infect Dis*. **2020**.
5. Centers for Medicare & Medicaid Services Technical Expert Panel Composition List. Project title: measure & instrument development and support (MIDS) patient safety development and maintenance: sepsis outcome measure. Available at: <https://www.cms.gov/files/document/patient-safety-measure-development-and-maintenance-sepsis-outcome-measure-tep-composition-list.pdf>. Accessed 5 February 2020.
6. Klompas M, Rhee C. The CMS sepsis mandate: right disease, wrong measure. *Ann Intern Med* **2016**; 165:517-8.
7. Rhee C, Kadri SS, Danner RL, et al. Diagnosing sepsis is subjective and highly variable: a survey of intensivists using case vignettes. *Crit Care* **2016**; 20:89.
8. 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.
9. Centers for Disease Control and Prevention. Hospital toolkit for adult sepsis surveillance. 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 23 March 2018.
10. 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**; 47:307-14.
11. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* **2016**; 315:801-10.
12. Rhee C, Jones TM, Hamad Y, et al; Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Netw Open* **2019**; 2:e187571.
13. Rhee C, Wang R, Song Y, et al. 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. *Crit Care Explor* **2019**; 1:e0049.
14. Seymour CW, Gesten F, Prescott HC, et al. Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med* **2017**; 376:2235-44.
15. 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**; 44:319-27.
16. Ash AS, Fienberg SE, Louis TA, et al. Statistical issues in assessing hospital performance. White Paper on Quantitative Health Sciences Publications and Presentations. **2012**. Available at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>. Accessed 12 April 2020.
17. Nelson JC, Pepe MS. Statistical description of interrater variability in ordinal ratings. *Stat Methods Med Res* **2000**; 9:475-96.
18. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg* **2018**; 126:1763-8.
19. Rhee C, Wang R, Jentzsch MS, et al. Comparison of hospital surgical site infection rates and rankings using claims versus National Healthcare Safety Network surveillance data. *Infect Control Hosp Epidemiol* **2019**; 40:208-10.
20. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* **2001**; 29:1303-10.
21. Iwashyna TJ, Odden A, Rohde J, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Med Care* **2014**; 52:e39-43.
22. Whittaker SA, Mikkelsen ME, Gaieski 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**; 41:945-53.
23. Walkey AJ, Shieh MS, Liu VX, Lindenauer PK. Mortality measures to profile hospital performance for patients with septic shock. *Crit Care Med* **2018**; 46:1247-54.
24. Wang HE, Donnelly JP, Shapiro NI, et al. Hospital variations in severe sepsis mortality. *Am J Med Qual* **2015**; 30:328-36.
25. Silva GC, Jiang L, Gutman R, et al. Mortality trends for veterans hospitalized with heart failure and pneumonia using claims-based vs clinical risk adjustment variables. *JAMA Intern Med* **2019**; 180(3):347-55.
26. Caroff DA, Wang R, Zhang Z, et al. The limited utility of ranking hospitals based on their colon surgery infection rates. [published online ahead of print January 9, 2020] *Clin Infect Dis*. ciaa012.
27. Darby JL, Davis BS, Barbash IJ, Kahn JM. An administrative model for benchmarking hospitals on their 30-day sepsis mortality. *BMC Health Serv Res* **2019**; 19:221.
28. 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**; 8:243-7.
29. Wunsch H, Rowan KM, Angus DC. International comparisons in critical care: a necessity and challenge. *Curr Opin Crit Care* **2007**; 13:725-31.
30. Krive J, Patel M, Gehm L, et al. The complexity and challenges of the *International Classification of Diseases, Ninth Revision, Clinical Modification* to *International Classification of Diseases, 10th Revision, Clinical Modification* transition in EDs. *Am J Emerg Med* **2015**; 33:713-8.
31. Caskey RN, Abutahoun A, Polick A, et al. Transition to *International Classification Of Disease Version 10, Clinical Modification*: the impact on internal medicine and internal medicine subspecialties. *BMC Health Serv Res* **2018**; 18:328.