# Supplementary Online Content

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*. doi:10.1001/jama.2017.13836

## I. eMethods, eTables, and eFigures



## II. eAppendices: Data Specifications for Implementing the EHR Sepsis Surveillance Definition



This supplementary material has been provided by the authors to give readers additional information about their work.

# I. SUPPLEMENTAL METHODS, TABLES, AND FIGURES

## eMethods 1. Healthcare Systems / Datasets Used in Study

The institutions involved in our study included a broad array of academic, community, and federal hospitals with diverse electronic health record (EHR) systems. For this study, acute care hospitals were included, while psychiatric, rehabilitation, and children's hospitals were excluded. Each dataset is described below.

1. **Hospital Corporation of America (HCA):** HCA is the largest for-profit healthcare system in the U.S., with 165 hospitals spread across 20 different states. The hospitals mainly consist of urban and suburban community medical centers but also include several complex tertiary referral and academic medical centers. From geographic, demographic, and socioeconomic standpoints, this population is representative of the U.S. population as a whole.[1] HCA hospitals primarily use the Meditech EHR system. Comprehensive clinical data is stored in a centralized data warehouse and undergoes line-item validation until >99% accuracy is achieved. This data warehouse has been used in the analysis of large cluster randomized trials, quality improvement initiatives, and other retrospective studies.[2-7] For this study, HCA provided complete data from 137 acute care hospitals for 2013 and 2014.
2. **Veterans Affairs (VA) Hospitals:** the VA is the largest integrated health system in the U.S and includes 130 hospitals, including at least one in every state. The VA has long been a national leader in the adoption of EHRs and has used a homegrown integrated EHR system since 1999.[8] Data from the VA EHR system has been used in multiple retrospective studies, including in the validation of the Sepsis-3 clinical criteria.[9-12] For this study, the VA provided complete data from 129 hospitals for 2014.
3. **University of Pittsburgh Medical Center (UPMC) System**: UPMC includes academic and community hospitals distributed throughout western Pennsylvania that collectively provide care for over 400,000 patients per year and utilize the Cerner EHR system. This dataset served as the primary dataset for derivation of the Sepsis-3 clinical criteria.[12] For this study, UPMC provided data from 11 hospitals for 2010-2014. The data did not overlap with the Cerner HealthFacts dataset.
4. **Cerner HealthFacts® Dataset:** rather than a single integrated healthcare system, HealthFacts is a de-identified patient database that receives granular clinical data from a subset of hospitals that use the Cerner EHR system. This dataset was first launched in 2000 and has been utilized in several retrospective studies spanning a range of different conditions.[13-19] For this study, we included 78 hospitals with complete data in 2014. A smaller subset of hospitals (at least 41 hospitals per year) provided data from 2009-2013.
5. **Institute for Health Metrics (IHM):** IHM is a non-profit organization that assists more than 100 hospitals across 30 states using the Meditech EHR system to extract and aggregate their electronic clinical data to support regulatory, reporting, quality improvement, and research initiatives.[20,21] The network primarily includes community hospitals with an average size of 160 beds. For this study, 49 hospitals provided one-half year of data for hospitalizations spanning July-December 2014. A subset of hospitals (at least 27 hospitals per year) provided data from 2009-2013.
6. **Brigham and Women's Hospital (BWH):** this academic medical center in Boston, Massachusetts provides inpatient care for over 50,000 patients per year. During the years of this study, BWH used a homegrown EHR system and stored detailed electronic clinical data from 2002 onwards in a clinical data warehouse (the Partners Research Patient Data Repository).[22] Data from this repository has been used for numerous research studies, including several studies of sepsis epidemiology and trends.[23-26]
7. **Emory Healthcare:** Emory Healthcare is the largest healthcare system in the state of Georgia and uses a Cerner EHR system. It has stored data in the Emory Clinical Data Warehouse since 1993 and has supported analysis of several quality improvement initiatives.[27,28] For this study, Emory Healthcare provided complete data from 4 hospitals for 2010-2014. The data did not overlap with the Cerner HealthFacts dataset.

## eMethods 2. Validation of EHR Surveillance Definition by Medical Record Reviews

In the Emory Healthcare system, one internist (R.D.) and one infectious disease physician (L.E.) reviewed a total of 310 hospitalizations using a standardized data abstraction tool in REDCap.[29] 110 cases were drawn from the surveillance definition-positive cohort, 110 were drawn from the definition-negative but presumed infection-positive cohort, and 90 were drawn from the definition-negative/infection-negative cohort. Reviewers were blinded to results of the EHR surveillance definition as well as discharge diagnosis codes. The initial 10 charts were reviewed together by the two reviewers; subsequently, 30 charts were reviewed independently and assessed for concordance. Agreement amongst those cases was good (kappa 0.73). The 4 cases where reviewers disagreed were excluded, yielding a total of 306 cases for analysis.

At Brigham and Women's Hospital (BWH), one intensivist (C.R.) reviewed 200 hospitalizations (100 surveillance definition-positive and 100 definition-negative, regardless of whether or not infection criteria was met), blinded to results of the surveillance definition and discharge diagnosis codes, using the same REDCap tool. This intensivist had previously been demonstrated to have good interrater reliability (kappa 0.80) compared to a second intensivist when adjudicating the presence of sepsis on medical record reviews in this dataset.[26]

Hospitalizations were classified as sepsis if there was either definite (i.e., positive cultures or radiography and compatible syndrome) or possible infection (i.e., infection presumed and treated by the medical team and no definitive alternate etiology found), and if organ dysfunction was present (with a rise in Sequential Organ Failure Assessment Score by ≥2 from baseline) and was felt to be related to infection. Baseline SOFA scores were determined by the reviewer based on an array of available data in the medical records, estimating the degree of organ dysfunction prior to infection onset (if infection occurred during hospitalization) or at baseline health (if infection was present-on-admission).

The EHR surveillance definition was then compared to the results of the medical record reviews. To calculate sensitivity, specificity, positive predictive value, and negative predictive value, the proportions were extrapolated back to the entire hospitalized population at the BWH and Emory hospitals. 10,000 bootstrapped samples were then created to calculate percentile-based confidence intervals for sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) using R version 3.3.1 (r-project.org). An identical method was used to calculate performance of claims-based definitions (explicit severe sepsis or septic shock codes, and the combination of implicit codes for infection and organ dysfunction [Angus method][30,31] or explicit severe sepsis / septic shock codes).

Sensitivity for EHR vs claims-based definitions was compared on the pooled number of sepsis-positive reviewed charts (n=163) using McNemar's test of paired proportions. The PPV for EHR vs claims-based definitions was compared by using a multinomial-based Wald statistic to test for equality. The test statistic in this paired study design was based on a bootstrapped estimate of the variance-covariance matrix, and accounts for correlations that arise from applying each surveillance definition to the same charts.[32] The results are summarized in **eTable 5**.

## eMethods 3. National Weighting Methodology for Sepsis Incidence and Mortality

In order to obtain a weighted estimate of sepsis incidence and mortality in all U.S. hospitals, study hospitals and American Hospital Association (AHA) acute care hospitals were stratified by bed size (large - ≥500 beds, medium - 200-499 beds, and small - <200 beds), teaching vs nonteaching status (as defined by the Accreditation Council for Graduate Medical Education), and U.S. region (Northeast, South, West, and Midwest). Hospitals in the AHA database were excluded if they were rehabilitation, pediatric, or psychiatric hospitals, or had zero general medical or surgical beds.

To account for the fact that the 49 Institute of Health Metrics (IHM) hospitals only had data from July-December 2014, sepsis incidence and death rates for January-June 2014 were imputed by averaging January-June sepsis incidence and mortality rates from IHM hospitals in 2013 and 2015. The total number of hospital admissions for January-June 2014 were also imputed by averaging the overall hospital admission rates in January-June vs July-December from calendar years 2013 and 2015.

The weighted sepsis incidence was obtained by a weighted mean of stratum specific sepsis incidence, estimated from study hospitals, with weights proportional to the number of AHA adult admissions in the stratum (**eTable 3** below). The number of adult admissions in each hospital was not available in the AHA database; thus, to project the number of adult sepsis cases in each stratum, we multiplied the ratio of adult to total (adult and pediatric) admissions in our study hospitals against the AHA estimate for total admissions in each stratum. The weighted mortality rate was calculated similarly. Confidence intervals were generated based on the variance.

Standard errors and confidence intervals for the estimates were generated based on the binomial distributions and accounting for both within- and between- hospital variations in sepsis and mortality rates.

**eTable 1. Characteristics of 2014 Study Hospitals versus American Hospital Association (AHA) Acute Care Hospitals**

| Hospital Characteristic | Study Hospitals (N=409) | | AHA Hospitals (N=4,810) | |
|---|---|---|---|---|
| | **N** | **% of Study Cohort** | **N** | **% of AHA Cohort** |
| Geographic Region | | | | |
| Northeast | 56 | 13.7% | 598 | 12.4% |
| South | 205 | 50.1% | 2,168 | 45.1% |
| Midwest | 60 | 14.7% | 1,096 | 22.8% |
| West | 88 | 21.5% | 948 | 19.7% |
| Teaching Status | | | | |
| Teaching | 152 | 37.2% | 1,425 | 29.6% |
| Nonteaching | 257 | 62.8% | 3,385 | 70.4% |
| AHA Hospital Size | | | | |
| Small (<200 beds) | 220 | 53.8% | 2,606 | 54.2% |
| Medium (200-499 beds) | 155 | 37.9% | 1,939 | 40.3% |
| Large (≥500 beds) | 34 | 8.3% | 265 | 5.5% |

## eTable 2. Hospitals Contributing Data Each Year for 2009-2014 Trends Analysis

| Year | BWH | Cerner | IHM | Emory | UPMC | HCA | VA[a] | Total | Total Adult Admissions[b] |
|---|---|---|---|---|---|---|---|---|---|
| 2009 | 1 | 50 | 33 | 0 | 0 | 0 | 0 | 84 | 696,807 |
| 2010 | 1 | 51 | 34 | 2 | 6 | 0 | 0 | 94 | 737,695 |
| 2011 | 1 | 48 | 27 | 2 | 9 | 0 | 0 | 87 | 747,236 |
| 2012 | 1 | 41 | 36 | 3 | 10 | 0 | 0 | 91 | 780,193 |
| 2013 | 1 | 73 | 53 | 3 | 11 | 137 | 0 | 278 | 2,485,637 |
| 2014 | 1 | 78 | 49 | 4 | 11 | 137 | 0 | 281 | 2,354,056 |

Abbreviations: BWH = Brigham and Women's Hospital, Cerner = Cerner HealthFacts, IHM = Institute of Health Metrics, Emory = Emory Healthcare, UPMC = University of Pittsburgh Medical Center, HCA = Hospital Corporation of America, VA = Veterans Affairs Hospitals

[a] Veterans Affairs Hospitals were not included for the trends analysis since they only provided data for 2014 and were unable to provide granular institution and hospital case mix data.

[b] The total number of adult hospitalizations from 2009-2014 was 7,801,624.

## eTable 3. Sepsis Stratified by Hospital Characteristics in Study vs AHA Hospitals in 2014

| Region | Size | Teaching Status | AHA Hospitals | Study Hospitals | AHA Admissions | Study Admissions | Study % of AHA Admissions | Study Adult Sepsis Cases | Study Adult Sepsis Deaths | Projected Sepsis Cases | Projected Sepsis Deaths |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MW | S | T | 42 | 9 | 93,330 | 41,769 | 44.8% | 2,488 | 385 | 5,559 | 860 |
| MW | M | T | 120 | 7 | 1,398,302 | 76,091 | 5.4% | 2,938 | 482 | 53,991 | 8,858 |
| MW | L | T | 36 | 3 | 947,093 | 43,155 | 4.6% | 2,774 | 406 | 60,879 | 8,910 |
| MW | S | N | 728 | 34 | 611,166 | 71,714 | 11.7% | 3,187 | 398 | 27,161 | 3,388 |
| MW | M+L[a] | N | 170 | 7 | 1,114,379 | 60,340 | 5.4% | 3,513 | 475 | 64,877 | 8,779 |
| NE | S | T | 30 | 8 | 93,136 | 35,237 | 37.8% | 1,582 | 247 | 4,182 | 652 |
| NE | M | T | 201 | 9 | 2,548,213 | 99,780 | 3.9% | 4,555 | 863 | 116,327 | 22,040 |
| NE | L | T | 56 | 5 | 1,944,739 | 119,804 | 6.2% | 8,123 | 1,610 | 131,858 | 26,135 |
| NE | S | N | 168 | 28 | 300,015 | 94,859 | 31.6% | 4,973 | 671 | 15,729 | 2,121 |
| NE | M | N | 138 | 4 | 989,448 | 45,689 | 4.6% | 2,241 | 319 | 48,525 | 6,905 |
| NE | L | N | 5 | 2 | 166,721 | 8,771 | 5.3% | 448 | 69 | 8,516 | 1,312 |
| S | S | T | 106 | 18 | 234,144 | 93,512 | 39.9% | 4,714 | 667 | 11,802 | 1,670 |
| S | M | T | 419 | 53 | 4,737,772 | 669,097 | 14.1% | 41,957 | 5,992 | 297,092 | 42,428 |
| S | L | T | 128 | 19 | 3,665,933 | 317,304 | 8.7% | 19,227 | 3,353 | 222,141 | 38,742 |
| S | S | N | 1,036 | 67 | 1,214,262 | 241,005 | 19.8% | 13,886 | 1,805 | 69,963 | 9,092 |
| S | M | N | 470 | 46 | 3,600,393 | 490,080 | 13.6% | 32,682 | 4,415 | 240,099 | 32,436 |
| S | L | N | 9 | 2 | 232,093 | 9,499 | 4.1% | 384 | 45 | 9,382 | 1,100 |
| W | S | T | 55 | 2 | 111,861 | 4,908 | 4.4% | 205 | 15 | 4,672 | 342 |
| W | M | T | 203 | 16 | 2,431,704 | 159,130 | 6.5% | 9,838 | 1,562 | 150,341 | 23,862 |
| W | L | T | 29 | 3 | 675,164 | 38,858 | 5.8% | 3,079 | 647 | 53,498 | 11,242 |
| W | S | N | 441 | 54 | 467,995 | 165,980 | 35.5% | 8,946 | 1,271 | 25,225 | 3,583 |
| W | M+L[a] | N | 220 | 13 | 1,764,184 | 149,071 | 8.4% | 10,037 | 1,523 | 118,784 | 18,023 |

Abbreviations: Regions - NE = northeast, MW = Midwest, S = south, W = west; Size - L = large (≥500 beds), M = medium (200-499 beds), S = Small (<200 beds); Teaching Status – N = nonteaching, T = teaching hospital. AHA = American Hospital Association.

[a] The Midwest Large Nonteaching hospital stratum was combined with the Midwest Medium Non-teaching hospital stratum because there was only 1 AHA hospital in that stratum. The West Large Nonteaching hospital stratum was combined with the West Medium Non-teaching hospital stratum for the same reason.

.

## eTable 4. Sepsis Mortality Rates per Age Group, Sex, Race, and Hospital Type in 2014

| Strata | Patients per Strata | Sepsis Cases | | Sepsis Deaths | |
|---|---|---|---|---|---|
| | **N** | **N** | **% of Strata** | **N** | **Strata Mortality %** |
| Overall | 2,901,019 | 173,690 | 100% | 26,061 | 15.0% |
| Age Category | | | | | |
| 20-39 years | 555,722 | 11,475 | 2.1% | 991 | 8.6% |
| 40-59 years | 731,904 | 40,975 | 5.6% | 5,228 | 12.8% |
| 60-79 years | 1,123,203 | 80,857 | 7.2% | 12,568 | 15.5% |
| ≥80 years | 490,190 | 40,383 | 8.2% | 7,274 | 18.0% |
| Sex[a] | | | | | |
| Male | 1,493,335 | 100,030 | 6.7% | 15,643 | 15.6% |
| Female | 1,407,617 | 73,660 | 5.2% | 10,418 | 14.1% |
| Race/Ethnicity[a] | | | | | |
| White | 1,946,801 | 117,081 | 6.0% | 17,396 | 14.9% |
| Black | 464,554 | 26,564 | 5.7% | 3,862 | 14.5% |
| Hispanic | 296,055 | 18,417 | 6.2% | 2,635 | 14.3% |
| Asian | 54,357 | 3,632 | 6.7% | 608 | 16.7% |
| Other | 79,210 | 4,497 | 5.7% | 832 | 18.5% |
| Hospital Teaching Status[b] | | | | | |
| Teaching | 1,660,773 | 98,516 | 5.9% | 15,708 | 15.9% |
| Nonteaching | 1,231,776 | 74,311 | 6.0% | 10,142 | 13.6% |
| Hospital Size[b] | | | | | |
| Small (<200 beds) | 670,070 | 35,599 | 5.3% | 4,841 | 13.6% |
| Medium (200-499 beds) | 1,697,047 | 104,431 | 6.2% | 15,153 | 14.5% |
| Large (≥500 beds) | 525,432 | 32,797 | 6.2% | 5,856 | 17.9% |

[a] Data was missing for 48 patients for sex and 60,042 patients for race/ethnicity.
[b] One hospital in Puerto Rico did not have data on hospital characteristics and was excluded from the analysis.

## eTable 5. Accuracy of EHR Definition vs Claims Relative to Medical Record Reviews

| | EHR Clinical Surveillance Criteria | Implicit or Explicit Codes | Explicit Sepsis Codes |
|---|---|---|---|
| **Sensitivity** | 69.7% [52.9, 92.0%] | 66.0% [51.4, 80.7%] | 32.3% [24.4, 43.0%] |
| **Specificity** | 98.1% [97.7, 98.5%] | 90.4% [89.6, 91.6%] | 99.3% [99.1, 99.5%] |
| **NPV** | 98.0% [95.9, 99.6%] | 97.6% [95.8, 98.8%] | 95.7% [93.7, 97.3%] |
| **PPV** | 70.4% [64.0, 76.8%] | 31.0% [24.9, 40.4%] | 75.2% [69.8, 80.6%] |

Abbreviations: NPV = negative predictive value, PPV = positive predictive value. Numbers in brackets refer to the 95% confidence intervals.

## eTable 6. EHR Sepsis Definition vs Sepsis-3 by Chart Reviews: False Negatives

| Reasons for EHR Surveillance False Negatives vs Sepsis-3 Chart Reviews | False Negatives (Total N=13) |
|---|---|
| Rise in SOFA score by ≥2 points but no EHR organ dysfunction[a] | **10** |
| Hypoxemia without requiring mechanical ventilation | 6 |
| Hypotension without requiring vasopressors | 2 |
| Abnormal Glasgow Coma Scale | 1 |
| SOFA by creatinine but no doubling in baseline creatinine | 3 |
| SOFA by bilirubin but not bilirubin >2.0 mg/dL with doubling from baseline | 1 |
| SOFA by thrombocytopenia but platelets not <100 cells/μL with >50% decrease from baseline | 2 |
| No blood cultures drawn | **2** |
| Rise in SOFA score by ≥2 points related to infection occurred outside of the +/-2 day EHR surveillance window around blood culture day[b] | **1** |

[a] Five cases had multiple contributors to a total rise in SOFA score by ≥2 points: one patient had hypoxemia not requiring mechanical ventilation (1 SOFA point) + hypotension not requiring vasopressors (1 SOFA point); one patient had hypoxemia not requiring mechanical ventilation (1 SOFA point) + abnormal Glasgow Coma Scale (1 SOFA point); two patients had abnormal creatinine and bilirubin contributing 1 SOFA point each; and one patient had abnormal creatinine and thrombocytopenia contributing 1 SOFA point each.

[b] The case that occurred outside the ±2 day surveillance window developed a rise in creatinine and hypoxemic respiratory failure leading to a rise in SOFA score by ≥2 points.

## eTable 7. EHR Sepsis Definition vs Sepsis-3 by Chart Reviews: False Positives

| Reasons for EHR Surveillance False Positives vs Sepsis-3 Chart Reviews | False Positives (Total N=57) |
|---|---|
| No infection present in retrospect despite blood culture order and antibiotics | **25** |
| Infection present, but EHR organ dysfunction not caused by infection:[a] | **20** |
| Vasopressor initiation | 12 |
| Mechanical ventilation initiation | 3 |
| Doubling in baseline serum creatinine or decrease in eGFR by ≥50% | 4 |
| Total bilirubin ≥2.0 mg/dL and doubling from baseline | 1 |
| Platelet count <100 cells/μL and decrease by ≥50% from baseline | 2 |
| EHR organ dysfunction criteria did not cause rise in SOFA by ≥ 2 points[b] | **12** |
| Lactate ≥2.0 mmol/L but no rise in SOFA by 2 | 11 |
| Total bilirubin ≥2.0 and doubling from baseline | 1 |
| Doubling in serum creatinine or decrease in eGFR by ≥50% | 1 |

[a] The various reasons why EHR organ dysfunction criteria were not caused by infection add up to more than the total N=20 in that category because 2 cases had multiple non-sepsis related organ dysfunctions. These included a patient who had a doubling in serum creatinine and mechanical ventilation unrelated to infection; and a patient with doubling in serum creatinine and decline in platelet count by >50% to <100 unrelated to infection.

[b] One patient had both lactate ≥2.0 and doubling in serum creatinine, without a corresponding rise in SOFA score by ≥2 points.

## eTable 8. Demographics and Clinical Characteristics for Sepsis Patients in 2014 Defined by Clinical Cultures

This table represents the patients who met EHR sepsis criteria when presumed infection was defined by any clinical culture test rather than blood cultures alone. The list of clinical cultures used for this definition is described in eAppendix C later in this Supplement.

| Characteristic | Patients with Sepsis by Clinical Cultures (n=204,663) |
|---|---|
| Age | |
| Mean (SD), years | 66.5 (15.5) |
| 20-39 years (No., %) | 13,622 (6.7) |
| 40-59 years (No., %) | 48,022 (23.5) |
| 60-79 years (No., %) | 95,251 (46.5) |
| ≥80 years (No., %) | 47,768 (23.3) |
| Sex | |
| Male (No., %) | 116,188 (56.8) |
| Female (No., %) | 88,474 (43.2) |
| Race/Ethnicity[a] | |
| White (No., %) | 138,385 (67.6) |
| Black (No., %) | 30,991 (15.1) |
| Hispanic (No., %) | 21,716 (10.6) |
| Asian (No., %) | 4,159 (2.0) |
| Other (No., %) | 5,249 (2.6) |
| Comorbidities | |
| Diabetes (No., %) | 72,356 (35.4) |
| Chronic Pulmonary Disease (No., %) | 62,552 (30.6) |
| Congestive Heart Failure (No., %) | 51,299 (25.1) |
| Renal Disease (No., %) | 54,475 (26.6) |
| Cancer (No., %) | 39,850 (19.5) |
| Dementia or Cerebrovascular Disease | 20,994 (10.3) |
| Liver Disease (No., %) | 20,416 (10.0) |
| HIV or AIDS (No., %) | 1,940 (0.9) |
| Clinical Characteristics | |
| Positive Blood Cultures[b] (No., %) | 26,385 (18.2) |
| Required ICU Admission (No., %) | 108,673 (53.1) |
| Median ICU LOS (Range)[c] | 5 (1-6) |
| Median Hospital LOS (Range) [c] | 10 (8-11) |
| Discharge Disposition[a] | |
| Home (No., %) | 103,503 (50.6) |
| In-Hospital Death (No., %) | 28,675 (14.0) |
| Hospice (No., %) | 12,032 (5.9) |
| Non-Acute Care Facility (No., %) | 50,213 (24.5) |
| Transfer to Acute Care Hospital (No., %) | 4,879 (2.4) |

Abbreviations: HIV=human immunodeficiency virus, AIDS = acquired immune deficiency syndrome, ICU = intensive care unit, LOS = length of stay

[a] Data was missing in 1 case for sex, 4,163 (2.0%) of cases for race/ethnicity and 5,361 (2.6%) of cases for discharge disposition.

[b] Blood culture results were available in 280 of the 409 hospitals in our datasets; the percentage of positive blood cultures reflects the denominator of sepsis cases in those 280 hospitals (n=170,735). Positive blood cultures excluded common skin contaminants and were counted if they occurred anytime during hospitalization. The number of patients with positive blood cultures is higher than that reported for the primary sepsis definition in Table 2 in the main manuscript because some patients met EHR clinical criteria for sepsis based on concurrent clinical cultures + antibiotics + organ dysfunction and had positive blood cultures at another point during hospitalization outside of the +/-2 day window where all criteria were required to occur.

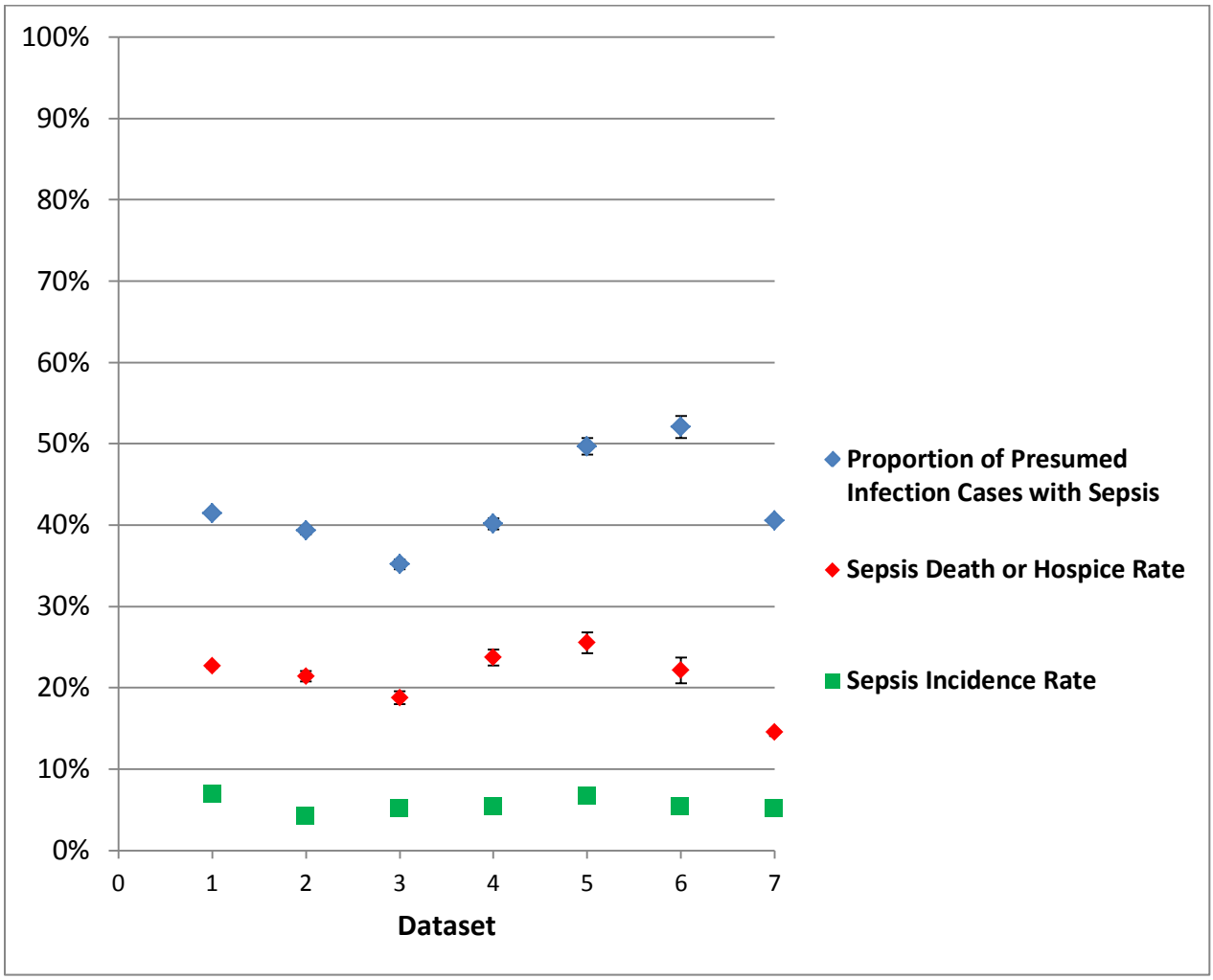
[c] The reported median ICU and hospital LOS represent the median of the median LOS values in each of the 7 datasets, along with the range of medians across the datasets.

## eTable 9. Trends for Sepsis Defined by Clinical Cultures

This table represents the adjusted trends in incidence, in-hospital mortality, and the combined outcome of in-hospital death or discharge to hospice, for patients who met EHR sepsis criteria when presumed infection was defined by any clinical culture test rather than blood cultures alone. "Sepsis With Lactate" refers to the surveillance definition that included elevate lactate as a criterion for organ dysfunction, while "Sepsis Without Lactate" excluded the lactate criterion. As with the trend analyses for the blood culture-based EHR surveillance definitions, VA hospitals were excluded.

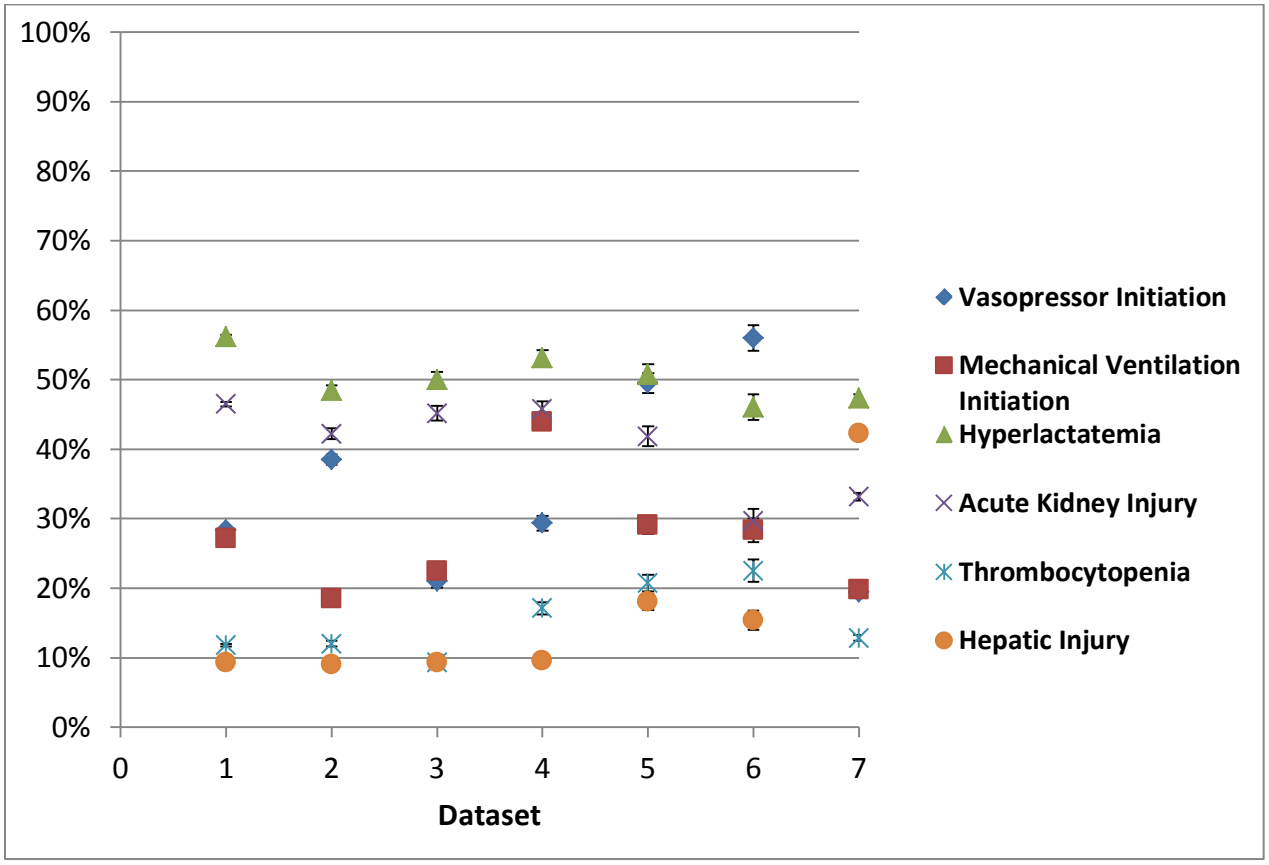
| | Sepsis With Lactate (by Clinical Cultures) | | | Sepsis Without Lactate (by Clinical Cultures) | | |
|---|---|---|---|---|---|---|
| | **Incidence** | **Mortality** | **Death or Hospice Discharge** | **Incidence** | **Mortality** | **Death or Hospice Discharge** |
| Adjusted 2009 vs 2014 rate | 6.1% [5.0-7.2%] vs 7.3% [7.2-7.3%] | 17.8% [15.8-19.8%] vs 14.1% [13.9-14.2%] | 23.2% [21.0-25.4%] vs 21.1% [20.9-21.3%] | 5.5% [4.4-6.6%] vs 5.8% [5.8-5.8%] | 18.9% [16.7-21.2%] vs 15.9% [15.6-16.1%] | 24.4% [21.9-26.9%] vs 22.6% [22.3-22.8%] |
| Annual Relative Change [95% CI] | +3.6% [-0.1,+7.3%] | -4.9% [-7.2,-2.5%] | -1.9% [-3.8,+0.0%] | +1.1% [-2.8,+5.1%] | -3.6% [-6.1,-1.2%] | -1.6% [-3.7,+0.5%] |
| p-value for trend | 0.06 | <0.001 | 0.052 | 0.58 | 0.003 | 0.14 |

## eFigure 1. Overall Sepsis Rates and Outcomes in Each Study Dataset in 2014




Each data point represents the mean rate of the specified variable in one of the 7 study datasets. The "presumed infection cases with sepsis" refers to the proportion of patients meeting presumed infection criteria (blood culture + 4 qualifying antibiotic days) that had ≥1 concurrent organ dysfunction criteria. Bars represent the 95% confidence intervals. The datasets are presented in no specific order.

## eFigure 2. Organ Dysfunction Rates Among Sepsis Cases in Each Study Dataset in 2014




Each data point represents the mean rate of the specified organ dysfunction criteria amongst all sepsis cases (defined using blood cultures) in one of the 7 study datasets. Bars represent the 95% confidence intervals. The datasets are presented in no specific order.

Thresholds for laboratory-based organ dysfunction criteria are as follows (as described in Box 1 in the main manuscript): hyperlactatemia - serum lactate ≥2.0 mmol/L; acute kidney injury - doubling in baseline creatinine or decrease in estimated glomerular filtration rate by ≥50%; thrombocytopenia - decrease in baseline platelets by ≥50% with baseline platelets >100 cells/μL; hepatic injury - doubling in baseline total bilirubin to ≥2.0 mg/dL.

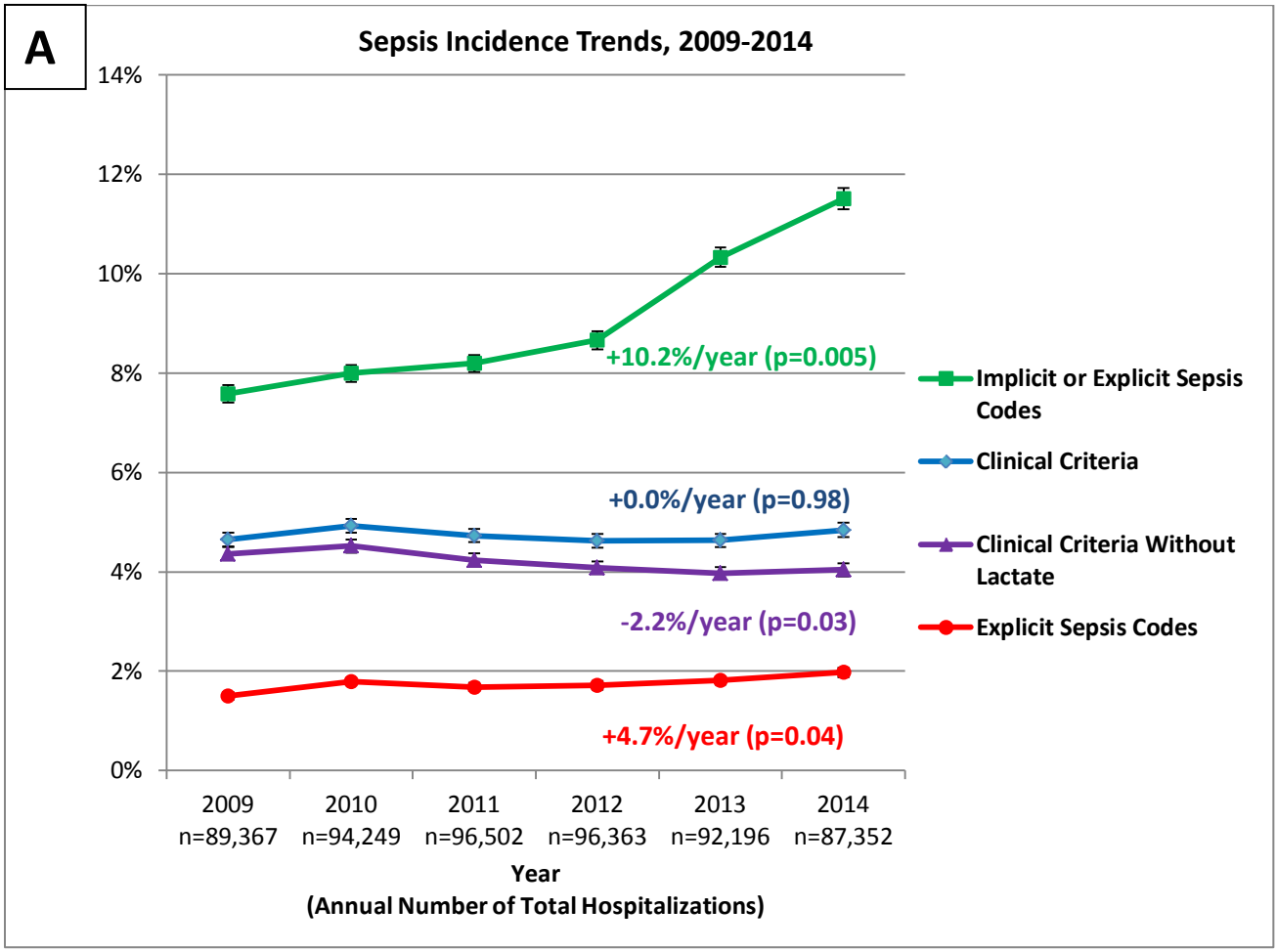
## eFigure 3. Sepsis Trends in Hospitals with Continuous Data from 2009-2014: A) Incidence, B) In-Hospital Mortality

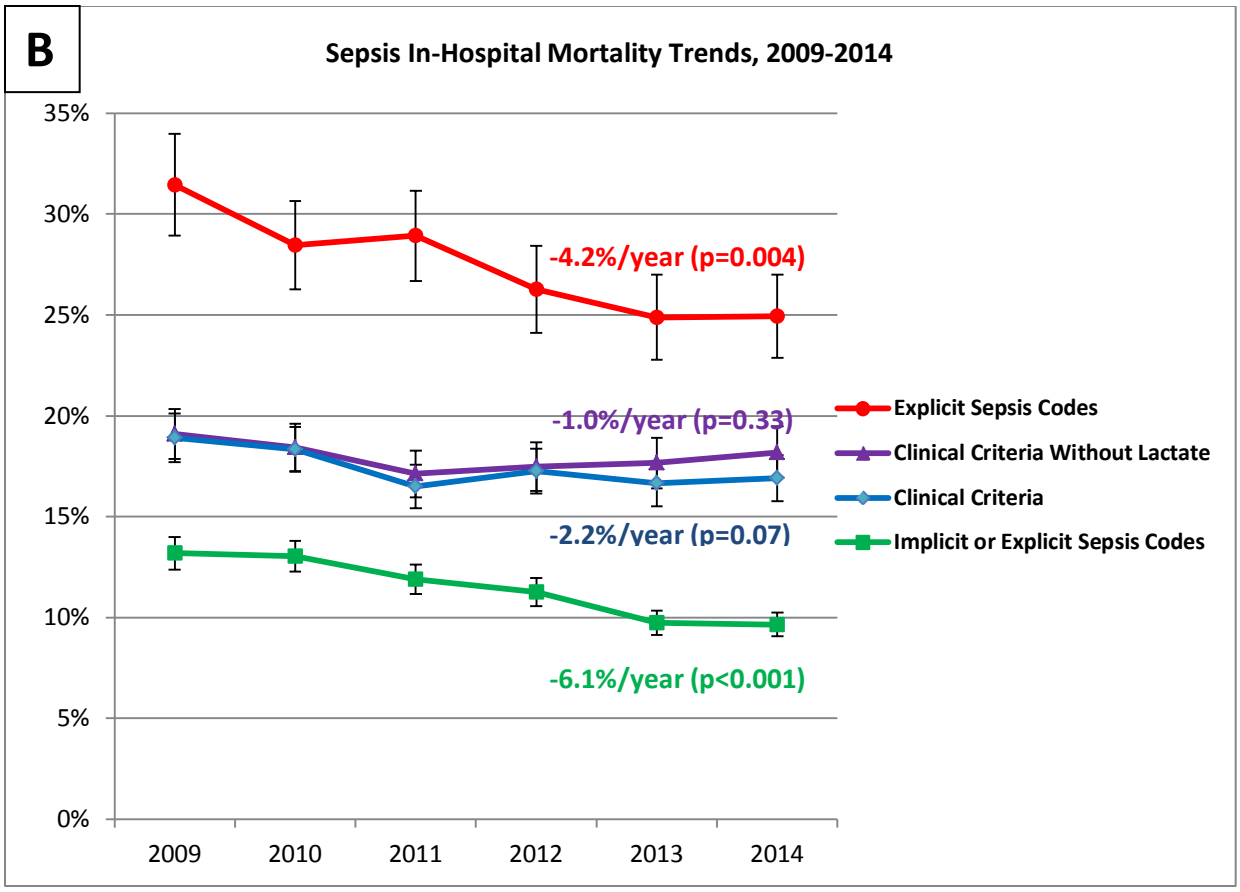
In this sensitivity analysis, there were 556,029 patients included at 6 hospitals that reported data each year from 2009-2014. Bars represent the 95% confidence intervals. Numbers adjacent to each line represent the mean relative change per year, with p-value indicative of significance of the trend.

"Clinical Criteria" refer to blood cultures + antibiotics and at least one concurrent organ dysfunction (refer to Box 1 in the main manuscript). "Clinical Criteria Without Lactate" is the same except without the criterion for lactate ≥2.0 mmol/L.

"Explicit Sepsis Codes" refers to hospitalizations with diagnoses of severe sepsis (995.92) or septic shock (785.52) on discharge. "Implicit Sepsis Codes" refers to hospitalizations with at least one infection diagnosis and one organ dysfunction diagnosis.

For panel B, the total number of sepsis cases per year for each definition was: Clinical Criteria – 4,154 (2009), 4,648 (2010), 4,565 (2011), 4,455 (2012), 4,270 (2013), 4,232 (2014); Clinical Criteria Without Lactate – 3,901 (2009), 4,261 (2010), 4,090 (2011), 3,934 (2012), 3,664 (2013), 3,531 (2014); Implicit or Explicit Codes – 6,779 (2009), 7,533 (2010), 7,908 (2011), 8,346 (2012), 9,526 (2013), 10,056 (2014); Explicit Codes – 1,342 (2009), 1,687 (2010), 1,622 (2011), 1,652 (2012), 1,672 (2013), 1,725 (2014).





## eReferences


1. 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(2):172-177.
2. Schweizer ML, Chiang HY, Septimus E, et al. Association of a bundled intervention with surgical site infections among patients undergoing cardiac, hip, or knee surgery. *JAMA.* 2015;313(21):2162-2171.
3. Huang SS, Septimus E, Kleinman K, et al. Targeted versus universal decolonization to prevent ICU infection. *N Engl J Med.* 2013;368(24):2255-2265.
4. Hayden MK, Lolans K, Haffenreffer K, et al. Chlorhexidine and Mupirocin Susceptibility of Methicillin-Resistant Staphylococcus aureus Isolates in the REDUCE-MRSA Trial. *J Clin Microbiol.* 2016.
5. Mah MP, Clark SL, Akhigbe E, et al. Reduction of severe hyperbilirubinemia after institution of predischarge bilirubin screening. *Pediatrics.* 2010;125(5):e1143-1148.
6. Septimus EJ, Hayden MK, Kleinman K, et al. Does chlorhexidine bathing in adult intensive care units reduce blood culture contamination? A pragmatic cluster-randomized trial. *Infect Control Hosp Epidemiol.* 2014;35 Suppl 3:S17-22.
7. Huang SS, Septimus E, Hayden MK, et al. Effect of body surface decolonisation on bacteriuria and candiduria in intensive care units: an analysis of a cluster-randomised trial. *Lancet Infect Dis.* 2016;16(1):70-79.
8. Evans DC, Nichol WP, Perlin JB. Effect of the implementation of an enterprise-wide Electronic Health Record on productivity in the Veterans Health Administration. *Health Econ Policy Law.* 2006;1(Pt 2):163-169.
9. Prescott HC, Kepreos KM, Wiitala WL, Iwashyna TJ. Temporal Changes in the Influence of Hospitals and Regional Healthcare Networks on Severe Sepsis Mortality. *Crit Care Med.* 2015;43(7):1368-1374.
10. Render ML, Deddens J, Freyberg R, et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. *Crit Care Med.* 2008;36(4):1031-1042.
11. Cooke CR, Kennedy EH, Wiitala WL, Almenoff PL, Sales AE, Iwashyna TJ. Despite variation in volume, Veterans Affairs hospitals show consistent outcomes among patients with non-postoperative mechanical ventilation. *Crit Care Med.* 2012;40(9):2569-2575.
12. 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(8):762-774.
13. Choudhry SA, Li J, Davis D, Erdmann C, Sikka R, Sutariya B. A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model. *Online J Public Health Inform.* 2013;5(2):219.
14. Goyal A, Spertus JA, Gosch K, et al. Serum potassium levels and mortality in acute myocardial infarction. *JAMA.* 2012;307(2):157-164.
15. Grodzinsky A, Goyal A, Gosch K, et al. Prevalence and Prognosis of Hyperkalemia in Patients with Acute Myocardial Infarction. *Am J Med.* 2016;129(8):858-865.
16. 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(8).
17. Chan WW, Waltman Johnson K, Friedman HS, Navaratnam P. Association between cardiac, renal, and hepatic biomarkers and outcomes in patients with acute heart failure. *Hosp Pract (1995).* 2016;44(3):138-145.
18. Andes D, Azie N, Yang H, et al. Drug-Drug Interaction Associated with Mold-Active Triazoles among Hospitalized Patients. *Antimicrob Agents Chemother.* 2016;60(6):3398-3406.
19. Petrick JL, Nguyen T, Cook MB. Temporal trends of esophageal disorders by age in the Cerner Health Facts database. *Ann Epidemiol.* 2016;26(2):151-154 e151-154.

20. Klompas M, Kleinman KP, Karcz A. Variability in mean duration of mechanical ventilation among community hospitals. *Infect Control Hosp Epidemiol.* 2012;33(6):635-637.
21. Howard DH, Karcz A, Roback JD. The accuracy of claims data for measuring transfusion rates. *Transfus Med.* 2016.
22. Murphy SN, Chueh HC. A security architecture for query tools used to access large biomedical databases. *Proc AMIA Symp.* 2002:552-556.
23. Rhee C, Murphy MV, Li L, et al. Comparison of Trends in Sepsis Incidence and Coding Using Administrative Claims Versus Objective Clinical Data. *Clin Infect Dis.* 2015;60(1):88-95.
24. Rhee C, Murphy MV, Li L, et al. Improving documentation and coding for acute organ dysfunction biases estimates of changing sepsis severity and burden: a retrospective study. *Crit Care.* 2015;19:338.
25. Rhee C, Murphy MV, Li L, et al. Lactate Testing in Suspected Sepsis: Trends and Predictors of Failure to Measure Levels. *Crit Care Med.* 2015;43(8):1669-1676.
26. Rhee C, Kadri S, Huang SS, et al. Objective Sepsis Surveillance Using Electronic Clinical Data. *Infect Control Hosp Epidemiol.* 2016;37(2):163-171.
27. Lyu PF, Hockenberry JM, Gaydos LM, Howard DH, Buchman TG, Murphy DJ. Impact of a Sequential Intervention on Albumin Utilization in Critical Care. *Crit Care Med.* 2016;44(7):1307-1313.
28. Murphy DJ, Lyu PF, Gregg SR, et al. Using Incentives to Improve Resource Utilization: A Quasi-Experimental Evaluation of an ICU Quality Improvement Program. *Crit Care Med.* 2016;44(1):162-170.
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381.
30. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001;29(7):1303-1310.
31. 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(6):e39-43.
32. Kosinski AS. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. *Stat Med.* 2013;32(6):964-977.

## II. DATA SPECIFICATIONS FOR IMPLEMENTING THE EHR SEPSIS DEFINITION

**Introduction / Overview**

The remainder of this document details the data elements and structure of input EHR datasets used to implement the case-finding sepsis surveillance definition and the analyses performed in the study. Essentially, after 9 input datasets were created, a series of SAS codes were run to produce a set of output data tables. The actual SAS code can be provided as a separate set of files by request to crhee@bwh.harvard.edu.

The project code anticipates the following files structure when running the analysis:

- Project Folder (e.g. "CDC National Sepsis Surveillance Project", specific name not necessary)
- Sub-Folder Name: "Raw Datasets" - includes all of the input datasets specified below
- Sub-Folder Name: "Working Datasets" - includes working datasets for SAS code
- Sub-Folder Name: "Output Tables" - includes results tables for final reporting
- Sub-Folder Name: "SAS Code" - includes all scripts unzipped

Step-by-step instructions are detailed below:

1. Ensure that all 9 input datasets are structured correctly for the analysis. These are detailed in **eAppendix D**, and are summarized briefly below:
    i. "Basic" – demographics and basic clinical characteristics (e.g. admit/discharge date, ICU admit/discharge date, discharge disposition)
    ii. "Diagnosis" – discharge diagnosis codes
    iii. "Procedure" – procedure codes (daily)
    iv. "Medication" – medication prescribing or dispensing information (on a daily basis), specifically antimicrobials and vasopressors*
    v. "Laboratory" – laboratory data (dates and results) for lactate, creatinine ± estimated GFR, total bilirubin, platelet count
    vi. "Blood_Culture" – date of blood culture draws (regardless of result)
    vii. "Pos_Blood_Culture" – date of positive blood cultures
    viii. "Clinical_Culture" – date of clinical culture draws (regardless of result)
    ix. "Hospital" – information on hospitals, if multiple hospitals in a dataset

    *The scripts assume that standardized medication names are included the Medication dataset before running the code. Please refer to **eAppendix B** below for information on the antimicrobials and vasopressors that were included. Cultures should be separated into three input datasets before running the code. These three datasets are Blood_Culture (blood cultures only), Pos_Blood_Culture (blood cultures with positive results), and Clinical_Culture (all clinical cultures) (described in **eAppendix C**).

2. Once the input datasets are prepared to match the documentation in the Data Spec, place the input datasets in an empty folder "Input Datasets".

3. Examine the table "Overview of SAS Programs" (**eAppendix E**) for a thorough explanation of each SAS program. This table includes the sequence, purpose, inputs, and outputs of each SAS program. A glossary of terms that are used in the output tables is provided in **eAppendix F**.

4. Begin running the SAS programs (provided separately). Start by running the file "0.0-user_inputs_v2.sas" to add the necessary information for the remaining SAS programs to run properly. This includes the full project path (ex: C:\Users\jdoe\Documents\…\CDC National Sepsis Surveillance Project) and the names of each input dataset placed into the "Input Datasets" folder. No additional changes beyond what is in this program should be needed to run the analysis. Then run the SAS programs in the order described in "Overview of SAS Programs", beginning with "0.0-user_inputs_v2.sas". As the programs run, they will fill the folder "Working Datasets" with interim datasets. The final six programs will fill the folder "Output Tables" with the results.

**eAppendix A. Detailed Description of EHR Sepsis Surveillance Definition**

PRESUMED SERIOUS INFECTION

**1.** Blood Culture obtained
- The blood culture day is the center of a +/-2 day surveillance window period. Sensitivity analyses were also performed using a broader set of clinical cultures (see eAppendix C).
- Multiple windows during a hospitalization are possible. The surveillance schedule utilizes a "sliding window" that allows for overlap if multiple blood cultures are drawn in short period of time.

**2.** ≥4 "Qualifying Antibiotic Days" (QAD) – starting within +/-2 days of BC Day:
- First QAD = first day in +/-2 day window period that patient receives a "new antibiotic" ("new" = not given in the prior 2 calendar days. "New" treats PO and IV formulations of an antibiotic the same way, EXCEPT for vancomycin). IV and IM antibiotics are treated as equivalents. See eAppendix B for list of antibiotics that count as potential QADs.
- Subsequent QADs can be the same antibiotic, or different antibiotics as long as first dose of each antibiotic in the sequence is "new" (i.e., not given in prior 2 calendar days).
- A gap of a single calendar day between administrations of the same antibiotic (PO or IV) count as QADs as long as the gap is not >1 day. (For example: IV Levofloxacin on Day 1, then PO Levofloxacin on Days 3, 5, and 7 = 7 QADs).
- There must be at least one new parenteral (IV/IM) antibiotic within the +/-2 day window period.
- For an inpatient, if death, discharge to another hospital, or discharge to hospice occurs before 4 QADs have elapsed, then the patient the patient can qualify for presumed infection with <4 QADs so long as they have consecutive QADs until day of, or 1 day prior to, death or discharge . For a patient only seen in the ED or ED observation unit, a QAD is required each day until day of, or 1 day prior to, death.

ACUTE ORGAN DYSFUNCTION

**3.** Any one of the following, within +/- 2 days of Blood Culture:
- Initiation of a new vasopressor (any dose for any duration): norepinephrine, dopamine, epinephrine, phenylephrine, vasopressin. To count as a new vasopressor, that specific vasopressor cannot have been administered in the prior calendar day.
- Initiation of invasive mechanical ventilation. To count as "initiation", there must be >1 calendar day between mechanical ventilation episodes. Invasive mechanical ventilation can be identified by ICD-9 codes (96.7, 96.71, 96.72) or CPT codes (94002, 94003, 94004, 94656, 94657), or directly from EHR clinical data if available.
- Doubling of serum creatinine or decrease by ≥50% of eGFR relative to baseline* - excluding patients with ICD-9 code for end-stage renal disease (585.6)
- Total bilirubin ≥ 2.0 mg/dL and doubling from baseline*
- Platelet count <100 cells/μL and ≥ 50% decline from baseline* - baseline must be≥100 cells/μL
- Serum Lactate ≥ 2.0 mmol/L (Note: for the primary analysis of sepsis trends over time, we excluded this criterion.)

Sepsis = Presumed Infection + ≥1 Acute Organ Dysfunction within +/-2 days of blood culture day
Septic Shock = Presumed Infection + Vasopressor + Lactate ≥2.0 mmol/L within ±2 days of blood culture day

***To determine the "baseline" laboratory values for Creatinine, eGFR, Total Bilirubin, and Platelets**:

**Determine if the patient had hospital-onset infection vs infection present-on-admission.**
- ***"Hospital-Onset Infection"*** (vs "Present-on-Admission" Infection) – requires that the Blood Culture Day and First QAD occur on or after day 3 of admission (where day 1 = day of admission). In this case, baseline creatinine or bilirubin = lowest value during +/-2 day window period (around the Blood Culture Day), and baseline eGFR and platelet count = highest value during +/-2 day window period.
- ***"Present-on-Admission Infection" –*** requires that the Blood Culture Day OR First QAD occur on day 1 or 2. In this case, baseline creatinine or bilirubin = lowest value during hospitalization, and baseline eGFR and platelet count = highest during hospitalization.

- If listed eGFR value is ≥60, treat eGFR as equal to 60
- Baseline platelet count for either scenario must be ≥100

"***Hospital Onset Sepsis" (vs Sepsis "Present-On-Admission") –*** requires all 3 of the following occur on calendar day 3 or later of hospitalization (where day 1 = admission day):

- Blood Culture, First QAD, and all Organ Dysfunction Criteria. In other words, the "Sepsis Onset Day" is the earliest day in the window period where the blood culture, first QAD, or organ dysfunction criteria is met. Hospital-onset sepsis requires sepsis onset day to be day 3 or greater.
- A sepsis case should be defined as **Present-on-Admission** Sepsis if any one of (Blood Culture Day, First QAD, or any Organ Dysfunction criteria) occurs on days 1 or 2.

When using blood culture, antibiotic, laboratory, or vasopressor / mechanical vent data to calculate sepsis cases, data from day -1 and day 0 (i.e. 2 days prior to admission) should be allowed to qualify for sepsis. (This is to account for data that might result from care in the emergency department prior to admission.)

**eAppendix B. Medications Used in Study (Antibiotics and Vasopressors)**

For purposes of case-finding criteria, the antibiotics were divided into parenteral (IV) and oral (PO) antibiotics. "Antibiotics" include antibacterial, antifungal, and antiviral agents. All PO and IV antibiotics were considered identical for purposes of determining whether an antibiotic is "new" or not (meaning that a switch from IV to PO or vice versa does not count as a "new antibiotic.") The one exception is IV vs PO Vancomycin (meaning that a switch from IV to PO vancomycin, or initiation of PO vancomycin while still on IV vancomycin, and vice versa, all count as "new" antibiotics). Intramuscular (IM) antibiotics were treated equivalently as IV antibiotics. The generic names of the antimicrobials included are listed below.

| ANTIBIOTICS |
|---|
| IV Antibacterials |
| amikacin, ampicillin, ampicillin/sulbactam, azithromycin, aztreonam, cefamandole, cefazolin, cefepime, cefmetazole, cefonicid, cefoperazone, cefotaxime, cefotetan, cefoxitin, ceftaroline, ceftazidime, ceftazidime/avibactam, ceftizoxime, ceftolozane/tazobactam, ceftriaxone, cefuroxime, cephalothin, cephapirin, chloramphenicol, ciprofloxacin, clindamycin, cloxacillin, colistin, dalbavancin, daptomycin, doripenem, doxycycline, ertapenem, gatifloxacin, gentamicin, imipenem, kanamycin, levofloxacin, lincomycin, linezolid, meropenem, methicillin, metronidazole, mezlocillin, minocycline, moxifloxacin, nafcillin, oritavancin, oxacillin, penicillin, piperacillin, pileracillin/tazobactam, polymyxin B, quinupristin/dalfopristin, streptomycin, tedizolid, telavancin, ticarcillin, ticarcillin/clavulanate, tigecycline, tobramycin, trimethoprim/sulfamethoxazole, vancomycin |
| PO Antibacterials |
| amoxicillin/clavulanate, amoxicillin, ampicillin, azithromycin, cefaclor, cefadroxil, cefdinir, cefditoren, cefixime, cefpodoxime, cefprozil, ceftibuten, cefuroxime, cephalexin, cephradine, chloramphenicol, cinoxacin, ciprofloxacin, clarithromycin, clindamycin, cloxacillin, dicloxacillin, doxycycline, fidaxomicin, fosfomycin, gatifloxacin, levofloxacin, lincomycin, linezolid, metronidazole, minocycline, moxifloxacin, nitrofurantoin, norfloxacin, ofloxacin, penicillin, pivampicillin, rifampin, sulfadiazine, sulfadiazine-trimethoprim, sulfamethoxazole, sulfisoxazole, tedizolid, telithromycin, tetracycline, trimethoprim, trimethoprim-sulfamethoxazole, vancomycin |
| IV Antifungals |
| amphotericin B, anidulafungin, caspofungin, fluconazole, itraconazole, micafungin, posaconazole, voriconazole |
| PO Antifungals |
| fluconazole, itraconazole, posaconazole, voriconazole |
| IV Antivirals |
| acyclovir, ganciclovir, cidofovir, foscarnet, peramivir |
| PO Antivirals |
| Oseltamivir |

The five vasopressors of interest were Norepinephrine, Epinephrine, Dopamine, Vasopressin, and Phenylephrine. We only included intravenous administrations of vasopressors, and excluded vasopressors that were clearly single bolus injections rather than continuous infusions.

**eAppendix C. Definitions of Blood Cultures and Clinical Cultures**

For the primary EHR surveillance definition, blood cultures were used as the center of the surveillance window (regardless of blood culture results). We included bacterial (aerobic and/or anaerobic), acid-fast bacilli (AFB), and fungal blood cultures. Blood cultures for specific viruses (e.g., cytomegalovirus) were excluded.

For a sensitivity analysis, a broader range of clinical cultures were used to define suspected infection, detailed below.

| TESTS INCLUDED AS "CLINICAL CULTURES" |
|---|
| Bacterial, fungal, or acid-fast bacilli cultures from the following sites: blood, urine, respiratory, cerebrospinal fluid, pleural, peritoneal, joint/synovial, abscess, wound, sinus, drain, catheter tip, medical devices, stool |
| Respiratory viral tests (rapid antigen, direct fluorescent antibody, polymerase chain reaction): influenza, parainfluenza, respiratory syncytial virus, adenovirus, metapneumovirus |
| Clostridium difficile toxin assays (ELISA or polymerase chain reaction) |
| Specific organism antigens from serum, urine, or cerebrospinal fluid, such as: Cryptococcus, Histoplasma, *Haemophilus influenza*, *Streptococcus pneumoniae*, *Legionella pneumophila* |
| Specific organism cultures or smears, such as: Pneumocystis, Nocardia, Legionella, Malaria |
| *Notable microbiology tests that were not included: Surveillance cultures (e.g., MRSA, VRE), tests for sexually transmitted diseases, HIV tests, parasite tests, H. pylori, Hepatitis, Fungal markers (galactomannan, beta-D-glucan), serological tests (IgM, IgG), Gram stains alone without a culture, non-respiratory viral cultures or PCRs (e.g., CMV)* |

Growth of any organism from a blood culture was considered to be positive, with the exception of the following common skin contaminants:

- Coagulase negative Staphylococci
- Bacillus species not anthracis
- Corynebacterium species
- Aerococcus species
- Micrococcus species
- Propionibacterium species

**eAppendix D. Input Datasets Required to Run Sepsis Algorithm**

**1. Dataset Name: "Basic"**

| Variable Name | Description | SAS Variable Type | Values |
|---|---|---|---|
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **hospital_id** | Unique Hospital Identifier | Numeric or Character | 000001, "000002",… 1, 2, 3... |
| **age_yrs** | Age in years at admission | Numeric | Integer Values |
| **gender** | Patient Gender | Numeric | 1 - M<br>2 - F |
| **race** | Patient Race-ethnicity<br>*Note: Hispanic is often listed as a race. If listed separately as Ethnicity, this takes precedence; . i.e. White Race + Hispanic Ethnicity = Hispanic* | Numeric | 1 - white<br>2 - black<br>3 - asian<br>4 - hispanic<br>5 - other<br>. - not identified / missing |
| **discharge_date** | Day of hospital discharge relative to admission day | Numeric | Integer Values |
| **discharge_dispo** | Discharge disposition of patient | Numeric | 1 - Home<br>2 - Transfer to another acute care hospital<br>3 - Subacute facility<br>4 - Death<br>5 - Hospice |
| **year** | Calendar year of hospitalization (based on date of admission) | Numeric | 4-digit year |
| **icu_admitdate** | Day of first admission to ICU (relative to admit date) | Numeric | Integer Values |
| **icu_los** | Total number of calendar days in the ICU | Numeric | Integer Values |

**2. Dataset Name: "Diagnosis"**

| Description: This table should include **one row for each unique diagnosis_code/day combination.** If a patient receives the same diagnosis code multiple times in a day, only a single row is needed. Repeated diagnosis codes in a single encounter but on different days receive separate rows. | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… |

| | | | |
|---|---|---|---|
| | | | 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **diagnosiscode** | Diagnosis code – either ICD-9 or CPT. No decimal points. | Character | ex: "99591" |
| **codetype** | Used to distinguish between ICD and CPT codes | Character | "ICD9" - ICD-9 codes "CPT4" - CPT codes |
| **sequenceno** | Sequence of codes at discharge (for ICD-9 codes). | Numeric | 1 - primary 2 - non-primary |
| **POA** | Present on admission flag (for ICD-9 codes) | Numeric | 1 - yes 0 - no |
| **primary_dx** | Primary (or Principal) Diagnosis at discharge. | Character | ex: "99591" |
| **admit_dx** | Admitting Diagnosis (if available) | Character | ex: "99591" |

**3. Dataset Name: "Procedure"**

Description: This table should include **one row for each unique procedure_code/day combination.** If a patient receives the same diagnosis code multiple times in a day, only a single row is needed. Repeated procedure codes in a single encounter but on different days receive separate rows.

| Variable Name | Description | SAS Variable Type | Values |
|---|---|---|---|
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **procedure_code** | Diagnosis code – either ICD-9 or CPT. No decimal points. | Character | ex: "9670" |
| **ICD9orCPT** | Used to distinguish between ICD and CPT codes | Character | "ICD9" - ICD-9 codes "CPT4" - CPT codes |
| **day** | Day on which the procedure occurred | Numeric | Integer Values |

**4. Dataset Name: "Medication"**

| Description: This table should include **one row for each unique medication administration (or medication prescription if administration is unavailable).** | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **day** | Day on which the medication was administered | Numeric | Integer Values |
| **med** | Medication name | Character | ex: "Vancomycin", "Cefepime" String case does not matter. |
| **Route** | Administration route of medication | Character | "IV" - Intravenous "PO" - Oral "Other" - Other Route |
| **med_type** | Type of medication provided | Numeric | 1: Vasopressors* 2: Systemic Antibiotics 3: Warfarin** 4: Inotropes** 5: Antifungal 6: Antiviral |

* Vasopressor medication type includes IV norepinephrine, epinephrine, phenylephrine, dopamine, and vasopressin. Systemic antibiotics include antibacterials, antifungals, and antivirals.
**Warfarin and inotropes were included in the original data spec but ultimately not used for the analysis.

**5. Dataset Name: "Laboratory"**

| Description: This table should include **one row for each day** that a patient received any of the relevant lab value measurements. | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **Day** | Day on which the lab values were measured | Numeric | Integer Values |
| **egfr_max*** | Maximum estimated glomerular filtration rate for day (mL/min/1.73m$^2$). Note that any eGFR value >60 should be set as equal to 60 | Numeric | Decimal Values |
| **egfr_min*** | Minimum estimated glomerular filtration rate for day (mL/min/1.73m$^2$). Note that any eGFR value >60 should be set as equal to 60 | Numeric | Decimal Values |
| **cr_max** | Maximum measured creatinine levels for day (mg/dL) | Numeric | Decimal Values |
| **cr_max** | Minimum measured creatinine levels for day (mg/dL) | Numeric | Decimal Values |
| **tbili_max** | Maximum measured bilirubin levels for day (mg/dL) | Numeric | Decimal Values |
| **tbili_min** | Minimum measured bilirubin levels for day (mg/dL) | Numeric | Decimal Values |
| **plt_max** | Maximum measured platelet levels for day ($10^9$/L) | Numeric | Decimal Values |
| **plt_min** | Minimum measured platelet levels for day ($10^9$/L) | Numeric | Decimal Values |
| **lactate_max** | Maximum measured lactate for day (mmol/L) | Numeric | Decimal Values |
| **lactate_min** | Minimum measured lactate for day (mmol/L) | Numeric | Decimal Values |

** Note that a lab reporting 'creatinine clearance' can be used interchangeably with eGFR*

**6. Dataset Name: "Blood_Culture"**

| Description: This table should include **all blood culture orders** for encounters in the basic dataset, where each row is a unique blood culture order. | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **bcx_drawn_day** | Day on which the blood culture was drawn relative to admission day | Numeric | Integer Values |

**7. Dataset Name: "Pos_Blood_Culture"**

| Description: This table should include **only positive blood culture orders** for encounters in the basic dataset, where each row is a unique blood culture order. "Positive blood culture" excludes common skin contaminants (e.g., coagulase negative Staphylococci, Bacillus sp. not anthracis, Corynebacterium, Aerococcus, Micrococcus, Propionibacterium sp.). | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **bcx_drawn_day** | Day on which the blood culture was drawn relative to admission day | Numeric | Integer Values |

**8. Dataset Name: Clinical_Culture**

| Description: This table should include **all blood culture orders and additional clinical cultures outlined in eAppendix C** for encounters in the basic dataset, where each row is a unique culture order. | | | |
|---|---|---|---|
| **Variable Name** | **Description** | **SAS Variable Type** | **Values** |
| **patient_id** | A unique ID for each patient, consistent across different hospitalizations | Numeric or Character | "000001", "000002",… 1, 2, 3... |

| | | | |
|---|---|---|---|
| **admission_id** | A unique identifier for every unique hospitalization. One patient with multiple admissions will have a unique admission_id for each admission. | Numeric or Character | 000001, "000002",… 1, 2, 3… |
| **bcx_drawn_day** | Day on which the clinical culture was drawn relative to admission day | Numeric | Integer Values |

**9. Dataset Name: "Hospital"**

Description: This table should include **one row for each hospital** represented in the dataset.

| Variable Name | Description | SAS Variable Type | Values |
|---|---|---|---|
| **hospital_id** | Unique Hospital Identifier | Numeric or Character | "000001", "000002",… 1, 2, 3... |
| **state_code** | 2-digit state abbreviation | Character | ex: "MA" ex: "TX" |
| **beds** | Total number of inpatient beds | Numeric | Integer Values |
| **teaching** | To denote if hospital teaching status | Numeric | 1 - Teaching hospital 2 - Nonteaching hospital |
| **annual_discharges** | Total number of discharges for patients aged 20+ per year | Numeric | Integer Values |

**eAppendix E. Overview of SAS Programs**

| Order | SAS code | Purpose | Input Dataset(s) | Output Dataset(s) |
|---|---|---|---|---|
| **0 - User Inputs** | | | | |
| 1 | 0.0-user_inputs_v# | Define project folder path and input dataset names | | |
| **1 - EHR SEPSIS DEFINITION** | | | | |
| **- Create Datasets** | | | | |
| 2 | 1.0-mechvent_v# | Calculate mechanical ventilator diagnosis using procedure data | raw.procedure | sepsis.mechvent |
| **- Sepsis Algorithm** | | | | |
| 3 | 1.1-ehr_infection_v# | Calculate infection cases using blood culture, medication, and basic dataset | raw.basic, raw.blood_culture, raw.medication | sepsis.ehr_infection |
| 4 | 1.2-ehr_organ_dysfunction_v# | Create organ dysfunction variables used in the sepsis definitions | raw.basic, raw.blood_culture, raw.diagnosis, raw.laboratory, raw.medication, sepsis.mechvent, sepsis.ehr_infection | sepsis.ehr_organ_dys |
| 5 | 1.3-ehr_sepsis_def_v# | Calculate sepsis variables using infection and organ dysfunction criteria | raw.basic, raw.blood_culture, sepsis.ehr_infection, sepsis.ehr_organ_dys | sepsis.ehr_sepsis_def |
| **2 - SENSITIVITY ANALYSIS** | | | | |
| **- Create Datasets** | | | | |
| 6 | 2.0-bacteremic_sepsis_v# | Create bacteremic sepsis dataset | raw.basic, raw.medication, raw.pos_blood_culture, sepsis.ehr_infection | sepsis.bacteremic_sepsis |
| **- Sepsis Algorithm** | | | | |
| 7 | 2.1-ehr_infection_cc_v# | Calculate infection cases using clinical culture, medication, and basic dataset | raw.basic, raw.medication, raw.clinical_culture | sepsis.ehr_infection_cc |
| 8 | 2.2-ehr_organ_dysfunction_cc_v# | Create organ dysfunction variables used in various sepsis definitions for clinical culture data | raw.basic, raw.diagnosis, raw.laboratory, raw.medication, raw.clinical_culture, sepsis.mechvent, sepsis.ehr_infection | sepsis.organ_dys_cc |

| | | | | |
|---|---|---|---|---|
| | | | _cc | |
| 9 | 2.3-ehr_sepsis_def_cc_v# | Calculate sepsis variables using infection and organ dysfunction criteria for clinical cultures | raw.basic, raw.clinical_culture sepsis.ehr_infection_cc, sepsis.ehr_organ_dys_cc | sepsis.ehr_sepsis_def_cc |
| 10 | 2.4-angus_v# | Macro code for angus dombrovskiy definitions | | |
| 11 | 2.5-ehr_angus_domb_v# | Calculate angus and dombrovskiy diagnosis | raw.basic, raw.diagnosis, raw.procedure | sepsis.ehr_angus_domb |
| 12 | 2.6-ehr_severe_sepsis_v# | Calculate severe sepsis definition | raw.diagnosis | sepsis.ehr_severe_sepsis |
| **3- CALCULATE COMORBIDITIES FOR FINAL TABLES** | | | | |
| **- Charlson Comorbidities** | | | | |
| 13 | 3.0-charlson_macro_v# | Macro code for Charlson Comorbidities | | |
| 14 | 3.1-sepsis_charlson_v# | Calculate Charlson Comorbidities using Charlson_macro | raw.basic, raw.diagnosis, raw.procedure | sepsis.charlson |
| **- Elixhauser comorbidities** | | | | |
| 15 | 3.2-comofmt_analy_v# | Create format library of ICD codes and labels for Elixhauser Comorbidities | | |
| 16 | 3.3-elixhauser_macro_v# | Macro code that creates comorbidity variables based on the presence of secondary diagnoses and redefines comorbidity group by eliminating DRGs directly related to them. | | |
| 17 | 3.4-elixhauser_v# | Calculate Elixhauser scores for comorbidities [NOTE: does not use DRG to calculate scores] | raw.basic, raw.diagnosis | sepsis.elixhauser |
| **4 - TABLES FOR FINAL REPORTING** | | | | |

| | | | | |
|---|---|---|---|---|
| 18 | 4.0-table1_sepsis_mortality1_v# | Table 1.1 Overall Incidence and Mortality of Adults (>= 20) meeting infection and all Sepsis definitions (stratified by hospital and community onset) criteria | raw.basic, raw.blood_culture, raw.hospital, sepsis.bacteremic_sepsis, sepsis.ehr_sepsis_def, sepsis.ehr_angus_domb, sepsis.ehr_severe_sepsis | table1_sepsis_mortality1.xml |
| 19 | 4.1-table1_sepsis_mortality2_v# | Table 1.2-4 Overall Incidence and Mortality of Adults (>= 20) meeting infection and all Sepsis definitions stratified by demographic criteria | raw.basic, raw.blood_culture, raw.hospital sepsis.bacteremic_sepsis, sepsis.ehr_sepsis_def, sepsis.ehr_angus_domb, sepsis.ehr_severe_sepsis | table1_sepsis_mortality2_agegrp.xml, table1_sepsis_mortality2_gender.xml, table1_sepsis_mortality2_race.xml |
| 20 | 4.2-table2_sepsis_v# | Table 2. Organ Dysfunction summary of Adult (>= 20) Patients Meeting Primary Sepsis Definition Stratified by ICU vs. Non-ICU Patients | raw.basic, sepsis.ehr_organ_dys, sepsis.ehr_sepsis_def | table2_sepsis.xml |
| 21 | 4.3-table3_sepsis_v# | Table 3. Characteristics of all Adult (>= 20) patients | raw.basic, raw.hospital, raw.pos_blood_culture | table3_sepsis.xml |
| 22 | 4.4-table4_sepsis_v# | Table 4.1 Characteristics of all Adult (>=20) patients for sepsis definitions | raw.basic, raw.hospital, raw.pos_blood_culture, sepsis.bacteremic_sepsis, sepsis.ehr_sepsis_def, sepsis.ehr_sepsis_def_cc, sepsis.ehr_angus_domb, sepsis.ehr_severe_sepsis, sepsis.elixhauser, sepsis.charlson | table4_sepsis1.xml, table4_sepsis2.xml, table4_sepsis3.xml, table4_sepsis4.xml, table4_sepsis5.xml, table4_sepsis6.xml, table4_sepsis7.xml, table4_sepsis8.xml, table4_sepsis9.xml, table4_sepsis10.xml, table4_sepsis11.xml, table4_sepsis12.xml, table4_sepsis13.xml, table4_sepsis14.xml |

| | | | | |
|---|---|---|---|---|
| 23 | 4.5-table5_sepsis_v# | Table 5. Characteristics of all Adult (>=20) patients for Angus, Dombrovskiy, Severe sepsis definitions | raw.basic, raw.hospital, raw.pos_blood_culture, sepsis.bacteremic_sepsis, sepsis.ehr_sepsis_def, sepsis.ehr_angus_domb, sepsis.ehr_severe_sepsis, sepsis.elixhauser, sepsis.charlson | table5_sepsis.xml |

**eAppendix F. Glossary of terms for output tables**

| Term | Description |
|---|---|
| "Hosp" | Hospital-onset (i.e., all criteria met on day ≥3 of hospitalization, where day 1 = day of admission) |
| "Comm" | Community-onset (i.e., any criteria met on day 1 or 2 of hospitalization)<br>*Note: Hospital-onset and community-onset designations are mutually exclusive. If a patient meets both community-onset and hospital-onset criteria during a single encounter, he/she will be classified as community-onset.* |
| "Hospcomm" | Either hospital-onset OR community-onset for the definition of interest. So, "hospcomm_sepsis" refers to cases meeting criteria for the primary definition. |
| "QAD" | Qualifying Antibiotic Day<br>*Note: Please refer to the formal Data Specification document for more information about the definition of QAD* |
| "Presumed Infection" | Encounters with a blood culture drawn and four consecutive QADs starting within a ±2 day window of the blood culture order day. Fewer than 4 QADs are also allowed if the patient dies or is discharged to hospice or an acute care hospital, with consecutive QADs until ≤1 day prior to death/discharge. |
| "Organ Dysfunction" | Any one of the following, within a ±2 day window of the blood culture order day<br>• Initiation of a new vasopressor<br>• Initiation of invasive mechanical ventilation<br>• Serum lactate ≥2.0 mg/dL<br>• Doubling of serum creatinine or decrease by ≥50% of eGFR relative to baseline<br>• Total bilirubin ≥ 2.0 mg/dL and increase by 100% from baseline<br>• Platelet count <100 cells/μL and ≥50% decline from baseline<br>*Note: Please refer to the formal Data Specification document for more information about these criteria (i.e. definition of baseline).* |
| **Sepsis Definitions** | **Description** |
| >= 1 Blood Culture | • At least one blood culture anytime during hospitalization |
| Infection | • At least one episode of presumed infection |
| Sepsis | *Primary sepsis definition*<br>• Presumed infection<br>• ≥1 concurrent organ dysfunction |
| Sepsis_nolactate | *Sepsis Without Lactate*<br>• Presumed infection<br>• ≥1 concurrent organ dysfunction, excluding lactate criteria |
| Shock1_sepsis | *Septic Shock Definition*<br>• Presumed infection<br>• Initiation of new vasopressor within ±2-day window of the blood culture order day<br>• Lactate ≥2.0 within ±2-day window of the blood culture order day<br>*Note: This is the definition that most closely matches the SCCM/ESICM's Sepsis-3 Septic Shock clinical definition.* |
| Angus | *Angus Claims-Based Sepsis Definition*<br>• Any of the following:<br>- ≥1 infection code and ≥1 organ dysfunction code<br>- Explicit severe sepsis (995.92)<br>- Septic shock (785.52) code |
| Severe Sepsis | *Simple Code Definition*<br>• Explicit severe sepsis code (995.92) or Septic shock code (785.52) |