

OPEN

Comparison of Automated Sepsis Identification Methods and Electronic Health Record–based Sepsis Phenotyping: Improving Case Identification Accuracy by Accounting for Confounding Comorbid Conditions

Katharine E. Henry, MSE¹; David N. Hager, MD, PhD²; Tiffany M. Osborn, MD, MPH³;
Albert W. Wu, MD⁴; Suchi Saria, PhD^{1,4,5}

¹Department of Computer Science, Johns Hopkins University, Baltimore, MD.

²Division of Pulmonary and Critical Care, Department of Medicine, Johns Hopkins University, Baltimore, MD.

³Division of Emergency Medicine, Department of Surgery, Washington University, St. Louis, MO.

⁴Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

⁵Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD.

Ms. Henry and Dr. Hager are co-first authors.

This work was performed at the Johns Hopkins University and Johns Hopkins Medicine.

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

Ms. Henry is entitled to royalties from a licensing agreement between Johns Hopkins University and Bayesian Health LLC. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflicts of interest policies. Dr. Hager discloses salary support and funding to his institution from the Marcus Foundation for the conduct of the Vitamin C, Thiamine, and Steroids in Sepsis Trial. Dr. Saria has grants from Gordon and Betty Moore Foundation, the National Science Foundation, the National Institutes of Health, the Defense Advanced Research Projects Agency, and the American Heart Association. She is a founder of and holds equity in Bayesian Health. She is also the scientific advisory board member for PatientPing. She has received honoraria for talks from a number of biotechnology, research, and healthtech companies. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflicts of interest policies. The remaining authors have disclosed that they do not have any potential conflicts of interest.

Supported, in part, by grants from Gordon and Betty Moore Foundation.

For information regarding this article, E-mail: ssaria@cs.jhu.edu

Copyright © 2019 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Crit Care Expl 2019;1:e0053

DOI: 10.1097/CCE.0000000000000053

Objective: To develop and evaluate a novel strategy that automates the retrospective identification of sepsis using electronic health record data.

Design: Retrospective cohort study of emergency department and in-hospital patient encounters from 2014 to 2018.

Setting: One community and two academic hospitals in Maryland.

Patients: All patients 18 years old or older presenting to the emergency department or admitted to any acute inpatient medical or surgical unit including patients discharged from the emergency department.

Interventions: None.

Measurements and Main Results: From the electronic health record, 233,252 emergency department and inpatient encounters were identified. Patient data were used to develop and validate electronic health record–based sepsis phenotyping, an adaptation of “the Centers for Disease Control Adult Sepsis Event toolkit” that accounts for comorbid conditions when identifying sepsis patients. The performance of this novel system was then compared with 1) physician case review and 2) three other commonly used strategies using metrics of sensitivity and precision relative to sepsis billing codes, termed “billing code sensitivity” and “billing code predictive value.” Physician review of electronic health record–based sepsis phenotyping identified cases confirmed 79% as having sepsis; 88% were confirmed or had a billing code for sepsis; and 99% were confirmed, had a billing code, or received at least 4 days of antibiotics. At comparable billing code sensitivity (0.91; 95% CI, 0.88–0.93), electronic health record–based sepsis phenotyping had a higher billing code predictive value (0.32; 95% CI, 0.30–0.34) than either the Centers for Medicare and Medicaid Services Sepsis Core Measure (SEP-1) definition or the Sepsis-3 consensus definition (0.12; 95% CI, 0.11–0.13; and 0.07; 95% CI, 0.07–0.08, respectively). When compared with electronic health record–based sepsis phenotyping, Adult Sepsis Event had a lower billing code sensitivity (0.75; 95% CI, 0.72–0.78) and similar

billing code predictive value (0.29; 95% CI, 0.26–0.31). Electronic health record–based sepsis phenotyping identified patients with higher in-hospital mortality and nearly one-half as many false-positive cases when compared with SEP-1 and Sepsis-3.

Conclusions: By accounting for comorbid conditions, electronic health record–based sepsis phenotyping exhibited better performance when compared with other automated definitions of sepsis.

Key Words: electronic data processing; electronic health records; informatics; machine learning; sepsis, severe sepsis

Following the widespread adoption of electronic health record (EHR) systems, there has been great interest in leveraging this technology to improve patient care (1–3). Because of its high prevalence, morbidity, and mortality, there has been particular interest in using EHR-based tools to study and improve patient care in sepsis (4–8). This syndrome, characterized by organ dysfunction resulting from a dysregulated host response to infection (9), accounts for 30–50% of all in-hospital deaths and nearly \$24 billion in spending annually in the United States (10–12). The retrospective identification of patients with sepsis is important for the purposes of examining epidemiologic trends, measuring the impact of best practice initiatives, and validating predictive alerts (4–6, 13–15). However, the quality of new informatics techniques to predict sepsis depends on the quality of sepsis identification. One of the challenges of identifying sepsis is that comorbidities can confound, or mimic the effects of, the symptoms of sepsis, leading to false identification and discrepancies between retrospective and real-time performance (16). To date, the most reliable indicator of sepsis has been clinical case review (13). However, this expensive and time-consuming process is not practical for large-scale initiatives and much debate remains about the best criteria to identify sepsis (17, 18).

Attempts to automate sepsis identification have included the use of billing codes (17, 19) and the implementation of deterministic criteria based on consensus definitions (9, 17, 19). Although billing codes are easy to extract and generally have high positive predictive value (PPV) for sepsis, they generally do not document the time of sepsis onset, and when compared with clinician review, they have low sensitivity (13, 20). They are also subject to the effects of variable provider practice patterns and reimbursement policies (21). Consensus-based definitions have been more consistent over time and across institutions, but rely on clinicians to confirm whether organ dysfunction is due to sepsis (9, 22) and therefore have low PPV when automated (13, 19).

Recently, the Centers for Disease Control created the Adult Sepsis Event (ASE) toolkit (19, 23). This strategy uses objective clinical criteria to more reliably identify sepsis cases (13, 19). ASE achieves a higher PPV than consensus-based definitions, in part, by implementing stricter criteria for indicators of organ dysfunction. Although ASE is more sensitive than billing codes at a similar PPV, it still misses approximately 30% of cases (13). We hypothesize that by expanding the indicators of organ dysfunction and leveraging informatics techniques designed for noisy EHR data, we can recognize and filter out the false, or confounding, signals

of acute and chronic comorbidities and therapeutic interventions to more reliably identify sepsis.

The purpose of this study is to provide a comparison of commonly used sepsis identification strategies and to introduce EHR-based sepsis phenotyping (ESP), a new retrospective approach that filters out markers of organ dysfunction attributable to comorbid conditions and therapeutic interventions before defining a patient as having sepsis.

METHODS

This study was approved by the Johns Hopkins Medicine Institutional Review Board (Protocol Numbers: NA_00092916 and IRB00112903).

Study Population

The analysis includes all medical and surgical patients 18 years old or older admitted to either Howard County General Hospital (HCGH, 285 beds; 2014–2018), Johns Hopkins Hospital (1154 beds; 2016–2018), or Johns Hopkins Bayview Medical Center (455 beds; 2016–2018). Only a patient's first encounter was included in the study. Data were electronically abstracted from each patient's EHR using the database management system Postgres structured query language (PostgreSQL version 9.6.9; Berkeley, CA). Extracted data included all vital signs, laboratory measurements, and therapeutic interventions during each patient's healthcare encounter. In addition, demographic and anthropometric data, billing codes, and medical history were collected. Patients were excluded if their encounter was not associated with at least one order for a complete blood count and one order for a basic metabolic panel. A complete description of the variables collected is provided in **Supplemental Table 1** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

Sepsis Definitions

Specific sepsis definition criteria are shown in **Table 1**.

Billing Codes. Cases of sepsis were identified based on the use of *International Classification of Diseases* (ICD) codes for severe sepsis (ICD, 9th Edition [ICD-9] code 955.92; ICD, 10th Edition [ICD-10] code R65.20) and septic shock (ICD-9 code 785.52; ICD-10 code R65.21) at any time during a patient's encounter. Cases of sepsis without organ dysfunction (ICD-9 code 995.51; ICD-10 code A41.9) are not considered in this analysis.

SEP-1. The Centers for Medicaid and Medicare Services Sepsis Core Measure (SEP-1) definition is based on the 2001 consensus definition of severe sepsis (22). For consistency with the other definitions presented, we will simply refer to this as “sepsis.” To automate this definition, the presence of an infection-related billing code is used to indicate suspicion of infection (11).

Sepsis-3. Sepsis-3 was automated as described in the 2016 consensus definition (9, 17). Baseline Sequential Organ Failure Assessment scores were assumed to be zero if no prior information was available. Suspicion of infection was indicated by the presence and timing of culture orders and antibiotics.

EHR Sepsis Phenotyping and the ASE Toolkit. ESP is an adaptation of ASE. As detailed by Rhee et al (19), ASE identifies suspicion of infection as the time of a blood culture provided that a

TABLE 1. Sepsis Definition by Infection and Organ Dysfunction Components

Definition	Suspicion of Infection	Acute Organ Dysfunction	Attribution to Sepsis
Billing codes	NA	NA	Presence of ICD, 9th Edition codes 995.52 or 785.52 or a corresponding ICD, 10th Edition code R65.20 or R65.21
Centers for Medicaid and Medicare Services Sepsis Core Measure (SEP-1)	At least two SIRS criteria met within 6 hr of documentation of suspected infection (indicated by the presence of an infection-related billing code)	Any of the following: <ul style="list-style-type: none"> • Creatinine > 2.0 mg/dL excluding patients with end-stage renal disease • Bilirubin > 2.0 mg/dL • Lactate > 2.0 mmol/L • Persistent hypotension (indicated by two hypotensive measurements within 6 hr) • Vasopressors • Invasive or noninvasive mechanical ventilation • Platelets < 100 cells/μL • INR > 1.5 • PTT > 60 	Organ dysfunction criteria met within 6 hr of suspicion of infection onset
Sepsis-3	Culture ordered within the 24 hr following antibiotic use or followed by antibiotics within 72 hr	Increase in baseline SOFA score of ≥ 2 points, change in renal SOFA score not included if patients had ESRD SOFA scores were computed as follows: <ul style="list-style-type: none"> • Respiratory system (1–4 points as the ratio of P_{aO_2} to F_{iO_2} decreases) • Nervous system (1–4 points as GCS decreases) • Cardiovascular system (1–4 points as mean arterial pressure decreases or vasopressor dependence increases) • Liver (1–4 points as bilirubin increases) • Coagulation (1–4 points as platelets decreases) • Kidneys (1–4 points as creatinine increases) 	Organ dysfunction criteria met within 48 hr prior or up to 24 hr after onset of suspicion of infection
Adult Sepsis Event	New IV antibiotic is initiated within 2 d of the blood culture (before or after) and that antibiotics are continued for at least 4 d or up until patient death or transfer to hospice or another acute care facility	Any of the following: <ul style="list-style-type: none"> • Initiation of vasopressors outside of surgery • Initiation of invasive mechanical ventilation • Serum lactate ≥ 2.0 mmol/L • Doubling of serum creatinine among patients without ESRD • Total bilirubin level ≥ 2.0 mg/dL and doubling from baseline • Platelet count < 100 cells/mL that is at least 50% less than baseline 	Organ dysfunction criteria met within 48 hr of blood culture
Electronic health record–based sepsis phenotyping	Either: <ul style="list-style-type: none"> • New IV antibiotic and at least 4 d of antibiotics or up until day of patient discharge to an acute care facility AND either culture within 48 hr OR documentation of sepsis • Documentation of sepsis and two SIRS criteria met within 6 hr of each other 	Any of the following in the absence of related comorbid conditions and treatments (Supplemental Table 2, Supplemental Digital Content 1, http://links.lww.com/CCX/A106): <ul style="list-style-type: none"> • Initiation of vasopressors • Initiation of any mechanical ventilation • Serum lactate ≥ 2.0 mmol/L • Doubling of serum creatinine among patients without ESRD • Total bilirubin level ≥ 2.0 mg/dL and doubling from baseline • Platelet count < 100 cells/mL that is at least 50% less than baseline • Persistent hypotension (systolic blood pressure < 90 mm Hg or mean arterial pressure < 65 mm Hg with a repeat blood pressure reading within 15 min) • GCS < 14 • INR > 1.5 or PTT > 60 	Organ dysfunction met within 48 hr of suspicion of infection onset (indicated by the earliest of the criteria met)

ESRD = end-stage renal disease, GCS = Glasgow Coma Scale, ICD = *International Classification of Diseases*, INR = international normalized ratio, NA = not applicable, PTT = partial thromboplastin time, SIRS = systemic inflammatory response syndrome, SOFA = Sequential Organ Failure Assessment.

new IV antibiotic is initiated within 2 days of the culture (before or after), and that antibiotics are continued for at least 4 days or up until patient death or discharge. In contrast, ESP does not require the presence of a blood culture if sepsis is explicitly documented in the EHR. This is because despite current guidelines, blood cultures are not always obtained (24). ESP also allows documentation of sepsis in the presence of two or more systemic inflammatory response syndrome (SIRS) criteria to be used as an indicator of suspicion of infection, taking SIRS onset as the time of suspicion of infection.

ESP uses the same markers of organ dysfunction as ASE, with the addition of persistent hypotension, decreased Glasgow Coma Score (GCS), and coagulopathy (Table 1). In addition, ESP is “trained” to recognize and dismiss causes of organ dysfunction (comorbid conditions) that could otherwise mimic sepsis (25). For example, the use of sedatives, presence of acute stroke, and indicators of a drug overdose are recognized alternate causes of an abnormal GCS. A similar list, or set of alternate causes, was developed for each indicator of organ dysfunction. To develop these sets, early versions of ESP were applied to data from one hospital (HCGH; September–December 2017). A subset of cases meeting the definition for sepsis was then reviewed to identify false positives and any alternate causes of organ dysfunction consistent with, but not due to sepsis. If the reviewer judged a cause to be systematic and well represented in the false-positive cases, this cause was added to the set of alternate cause of organ dysfunction. The process was repeated until case review did not yield any new causes. A complete list of these indicators is provided in **Supplemental Table 2** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

Physician Review of ESP-Labeled Cases

Because ESP is new and has not been previously validated, we first present a comparison of this definition to physician chart review (HCGH; January–December 2018). In order to collect a large number of reviews, we approximated ESP to surface cases to physicians to review in real time, allowing them to review patient charts with the full clinical context and thus streamlining the review process. The approximation was done by removing the duration requirements for determining suspicion of infection. We later removed cases that did not meet the full retrospective ESP definition.

The review consisted of asking the physician to evaluate whether identified patients were suspected of having an infection, and when organ dysfunction was present, if it was attributable to the infection. This process is further described in **Supplemental Figure 1** (Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>). For patients identified by ESP, we report the PPV, the proportion of identified cases confirmed as having sepsis by the clinician reviewer. Because review was conducted in real-time, it is possible that some unconfirmed cases may have later evolved. As such, we also report the proportion that received more than 4 consecutive days of antibiotics is reported as a surrogate marker indicating significant concern for infection. We also report the proportion that ultimately had a sepsis-related billing code, either with organ dysfunction (ICD-9 codes 995.92 or 785.52 or ICD-10 codes R65.20 or R65.21) or without (ICD-9 code 995.91 or ICD-10 code A41.9). Unlike when computing billing code sensitivity

(BCS) and billing code predictive value (BCPV), as described below, here, we include ICD-9 code 995.91 and ICD-10 code A41.9 because, in practice, some providers use that code even when organ dysfunction is present.

Comparison of ESP to Other Automated Definitions

In the absence of a gold-standard definition for sepsis and the impractical nature of clinician chart review, the performance of each definition was next compared with the billing code definition. Then, the patients identified by each definition were compared base on 1) their in-hospital mortality rate; 2) the proportion with an ICU length of stay (LOS) greater than or equal to 3 days; and 3) the prevalence of organ dysfunction due to conditions other than sepsis.

Comparison to the Billing Code Definition. Although billing codes generally lack sensitivity, clinical case review had shown them to exhibit high specificity and PPV (13). A highly sensitive definition should, therefore, identify most cases associated with a sepsis billing code in addition to many other cases that do not have a billing code for sepsis. We refer to the fraction of cases with a billing code identified by a given method as BCS and the fraction of cases identified by a given method that have a billing code as the BCPV. When comparing two definitions, at similar BCS, higher BCPV is associated with lower false-positive rates. Because billing codes have low sensitivity compared with clinical case review, the BCPV provides a lower bound on a definition's PPV. To compute 95% CIs for BCS and BCPV, the bootstrap method was used (Python version 2.7.13; Wilmington, DE). Outcomes for each definition were recomputed from 10,000 patient encounters randomly sampled with replacement from those identified as having sepsis by that definition. This process was then repeated 1000 times, and the CIs were estimated.

In-Hospital Mortality and ICU LOS Greater Than or Equal to 3 Days. The proportion of patients identified by each definition who died, and the proportion with an ICU LOS greater than or equal to 3 days, was calculated. Although mortality and ICU LOS are not synonymous with a diagnosis of sepsis, they are objective and easily captured endpoints that are more likely to occur inpatient with sepsis than those with simple infections (9).

Organ Dysfunction Not Due to Sepsis. The prevalence of conditions other than sepsis causing organ dysfunction was computed for each definition. These included myocardial infarction, stroke, gastrointestinal bleeding, end-stage renal disease with dialysis, surgery with general anesthesia, and coronary artery bypass (CABG) surgery. Definitions of sepsis with high false-positive rates will more often exhibit these conditions. The criteria used to identify each of these conditions are detailed in Supplement Table 1 (Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

RESULTS

Population characteristics were similar across hospitals (**Supplemental Table 3**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>). Missing data necessary for the application of the different sepsis definitions were infrequent (**Supplemental Table 4**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

Physician Review of ESP-Labeled Cases

Of the encounters at HCGH from January to December 2018, 1,012 cases were surfaced and reviewed by a physician. Of those, 781 (77%) were retrospectively identified by ESP as having sepsis and included in the analysis. ESP had a PPV of 79%, with 619 identified patients confirmed in real-time by a physician as having sepsis (Fig. 1). Of the patients who were not confirmed at the time of the review, many were later coded as having sepsis or received over 4 consecutive days of antibiotics. Overall, 710 encounters (88%) were confirmed or coded as having sepsis and 777 (99%) were confirmed, coded, or received at over 4 consecutive days of antibiotics or were on antibiotics up until they were transferred to another acute care facility or died in-hospital.

Comparison to the Billing Code Definition

Of 233,252 included patient admissions, 0.9% of the overall population had a sepsis billing code, which corresponded to 2,050 unique patients across all three hospitals (Table 2). With a similar BCS to SEP-1 and Sepsis-3, ESP had a significantly higher BCPV. Although ASE also had a high BCPV, it exhibited a significantly lower BCS than the other methods. Results were similar across the three hospitals (Supplemental Table 5, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

Mortality and ICU LOS Greater Than or Equal to 3 Days

In-hospital mortality was lowest among patients identified by SEP-1 (7%) and Sepsis-3 (6%) and highest among patients

identified by billing codes (26%), with mortality among those identified by ASE and ESP in the middle (17% and 16%, respectively) (Table 3; and Supplemental Table 5, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>). A similar trend was observed for the outcome of ICU LOS of greater than or equal to 3 days (Table 3; and Supplemental Table 5, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>).

Organ Dysfunction Not Due to Sepsis

Patients with organ dysfunction not related to infection were more often identified by SEP-1 and Sepsis-3 than by ASE and ESP (Table 4). For instance, SEP-1 and Sepsis-3 identified 30% and 39%, respectively, of patients who had a myocardial infarction as having sepsis, whereas ASE and ESP both identified only 13% of patients and billing codes included 6% of patients. Similar trends were seen for patients with stroke and gastrointestinal bleeding. Additionally, SEP-1 and Sepsis-3 identified many more patients who underwent a general surgical procedure as having sepsis than did ASE or ESP. Among patients who had CABG surgery, nearly one third were identified as having sepsis by SEP-1, and over half were identified by Sepsis-3. In contrast, less than 1% of these patients were identified by ASE or ESP.

DISCUSSION

This article compares four commonly used methods to automate sepsis identification and proposes a new automated sepsis definition, ESP, that leverages the richness of data in the EHR to filter out

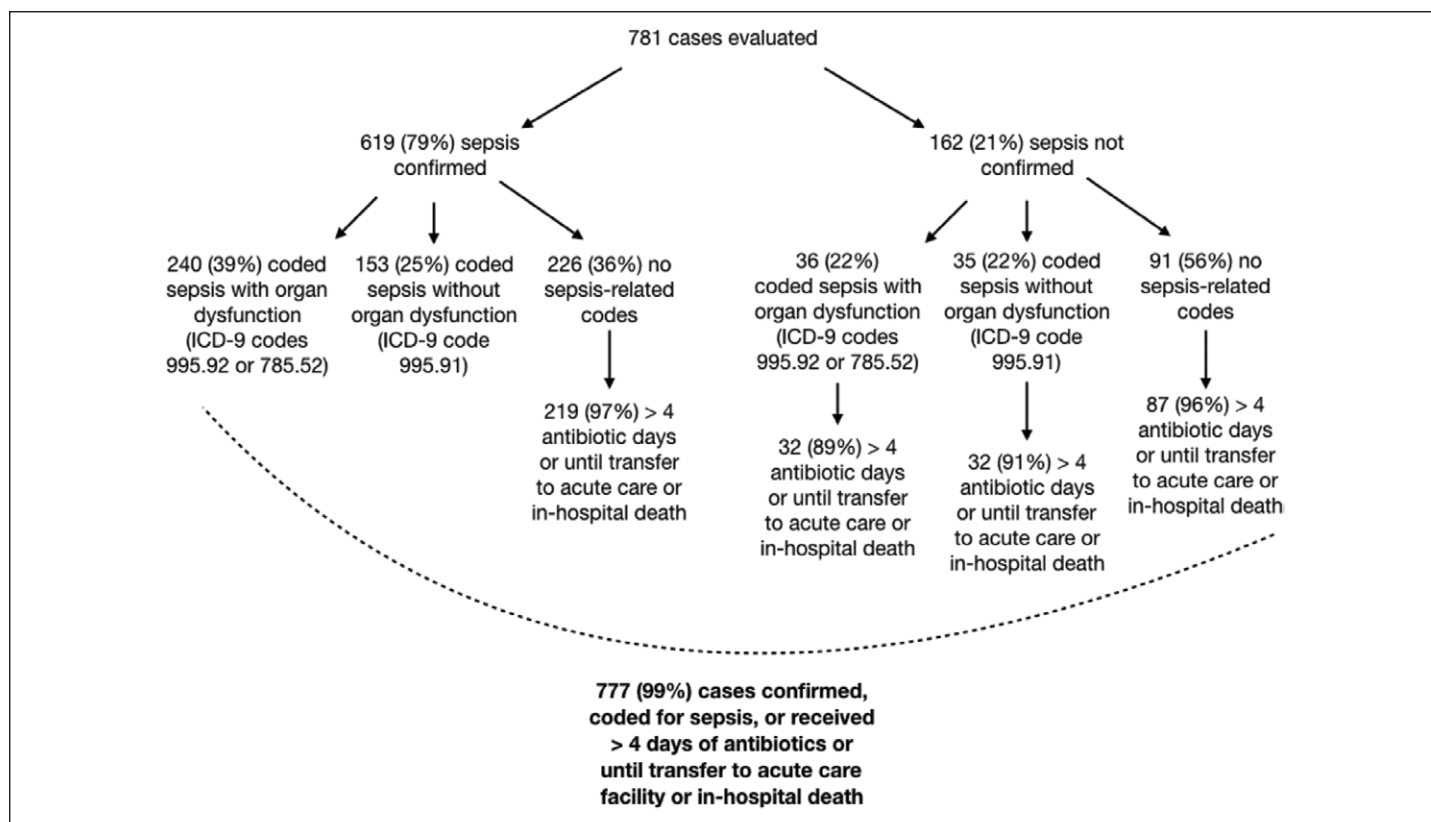


Figure 1. Real-time physician evaluation of electronic health record-based sepsis phenotyping (ESP)-identified cases ($n = 781$). The flowchart shows the breakdown of the subset of ESP-identified sepsis cases with physician review based on physician confirmation of sepsis, duration of antibiotics, and sepsis billing codes. ICD-9 = *International Classification of Diseases*, 9th Edition.

TABLE 2. Sensitivity and Predictive Value of Different Definitions of Sepsis Using Billing Codes as a Point of Reference

Definition	Total Encounters Identified by the Definition	Encounters With Billing Code	Billing Code Sensitivity (95% CI)	Billing Code Predictive Value (95% CI)
Centers for Medicaid and Medicare Services Sepsis Core Measure (SEP-1)	16,238	1,949	0.95 (0.92–0.96)	0.12 (0.11–0.13)
Sepsis-3	26,259	1,961	0.96 (0.94–0.97)	0.07 (0.07–0.08)
Adult Sepsis Event	5,382	1,540	0.75 (0.72–0.78)	0.29 (0.26–0.31)
Electronic health record–based sepsis phenotyping	5,817	1,858	0.91 (0.88–0.93)	0.32 (0.30–0.34)

TABLE 3. In-Hospital Mortality and Number of Patients With at least a 3-Day ICU Length of Stay by Sepsis Identification Method

Definition	Total Encounters Identified by the Definition	Deaths (%) ^a	ICU Length of Stay ≥ 3 d (%)
Billing code	2,050	528 (26)	1,084 (53)
Centers for Medicaid and Medicare Services Sepsis Core Measure (SEP-1)	16,238	1,150 (7)	3,634 (22)
Sepsis-3	26,259	1,492 (6)	5,661 (22)
Adult Sepsis Event	5,382	895 (17)	2,221 (41)
Electronic health record–based sepsis phenotyping	5,817	918 (16)	2,224 (38)

^aIn-hospital.**TABLE 4. Patients With Select Comorbidities Identified as Having Sepsis**

Definition	Gastrointestinal Bleeding <i>n</i> = 581 (%)	Myocardial Infarction <i>n</i> = 5,596 (%)	Stroke <i>n</i> = 5,229 (%)	End-Stage Renal Disease <i>n</i> = 3,157 (%)	Surgery With General Anesthesia ^a <i>n</i> = 26,380 (%)	Coronary Artery Bypass Graft <i>n</i> = 377 (%)	Any <i>n</i> = 38,463 (%)
Billing code	34 (6)	318 (6)	100 (2)	177 (6)	513 (2)	4 (1)	939 (2)
Centers for Medicaid and Medicare Services Sepsis Core Measure (SEP-1) (BCS 0.95, BCPV 0.12)	135 (23)	1,665 (30)	752 (14)	878 (28)	2,074 (8)	121 (32)	4,882 (13)
Sepsis-3 (BCS 0.96, BCPV 0.07)	167 (29)	2,186 (39)	1,364 (26)	1,115 (35)	4,088 (16)	199 (53)	8,095 (21)
Adult Sepsis Event (BCS 0.76, BCPV 0.28)	67 (12)	705 (13)	328 (6)	296 (9)	299 (1)	1 (0)	1,462 (4)
Electronic health record–based sepsis phenotyping (BCS 0.91, BCPV 0.32)	82 (14)	742 (13)	313 (6)	348 (11)	474 (2)	1 (0)	1,680 (4)

BCPV = billing code predictive value, BCS = billing code sensitivity.

^aSepsis onset (as indicated by each method) was within 6 hr prior and up to 24 hr after the surgical procedure except in the case of billing codes because billing codes do not have an associated time of onset.

confounding comorbidities when identifying sepsis cases. Definitions were compared on the basis of how well they predicted criteria associated with sepsis (sepsis billing codes, in-hospital mortality, and ICU LOS ≥ 3 d), as well as how often patients with potentially confounding comorbidities were identified by the methods. Overall, ESP achieved over twice the precision of SEP-1 and Sepsis-3 at the same sensitivity

of SEP-1, as measured by BCS and BCPV, and had higher rates of sepsis-associated outcomes like in-hospital mortality and extended ICU stay (Table 2). Because ESP has not previously been validated, we also compared it to physician review that was conducted in real time. ESP achieved a PPV of 79% when compared with physician review and 99% when compared with the combined outcome of being confirmed

by a physician, receiving a sepsis billing code, or having over 4 consecutive days of antibiotics (Fig. 1).

SEP-1 and Sepsis-3 were primarily developed for clinical use and were not intended to be automated for retrospective case identification (9, 22). They both require physician review to confirm organ dysfunction is likely related to sepsis, and in the case of SEP-1, to indicate suspicion of infection. As a result, when automated, these methods have high false-positive rates (Table 2) and often misattribute changes in signals related to other comorbidities to sepsis (26, 27). For instance, in our analysis, they identified many more patients undergoing surgical procedures as having sepsis than ASE or ESP. This was in large part because many of the patients undergoing procedures also received anesthesia and were on mechanical ventilation and were cultured and received antibiotics at some point after the surgical procedure. Although SEP-1 and Sepsis-3 specify that a clinician should exclude these cases, without a reliable way of automating this exclusion process, studies automatically implementing these definitions will routinely identify such patients as having sepsis.

ASE overcomes the low precision of automated implementations of SEP-1 and Sepsis-3 by using stricter criteria for change in organ dysfunction and infection. For instance, ASE requires the presence of a blood culture and over 4 consecutive days of antibiotics rather than any culture and at least one dose of antibiotics. However, in doing so, ASE significantly decreases its sensitivity compared with other methods and misses 21% of cases with a sepsis billing code that were identified by SEP-1 and Sepsis-3 (Table 2).

ESP was designed to maintain the high precision of ASE, while also increasing its sensitivity. It achieves this by using more expansive criteria for suspicion of infection and acute organ dysfunction, such as any culture type and persistent hypotension, although simultaneously leveraging the richness of data in the EHR to rule out indicators of organ dysfunction that are due to confounding comorbid conditions. For instance, a change in GCS was not used as an indicator of organ dysfunction if the patient was known to be sedated. The improved performance characteristics of this retrospective tool are important when considering the relative value of an intervention on patients with sepsis across a large healthcare network. Missed cases and false positives will lead to a noisier and less accurate understanding of cause and effect.

A natural question that arises is, given the number of sepsis definitions that already exist in clinical care (9, 22, 23), why is there a need for another one. Although multiple definitions exist, none to date is able to meet all of the priorities of different stakeholders (18, 28). Instead they attempt to tradeoff between different priorities depending on the use case, often prioritizing low implementation or measurement burden over reliability. However, the increased availability of the EHR combined with the development of informatics techniques presents new opportunities to achieve high reliability while also having low costs of implementation. New tools can pull data directly from the EHR and automate the implementation of the different criteria in coded queries. Critically, once developed, these queries can quickly and easily be repeated on new data. The up-front investment in time to develop the code to query an EHR is well offset by the potential to improve quality assurance and quality improvement programs. By leveraging more complex sources of information in the EHR, ESP was able to achieve higher combined precision and sensitivity than

previous methods. Furthermore, although ESP is intended for retrospective review using the patient's full chart, the cases it identifies could be used to develop real-time predictive methods that provide timely sepsis identification for use at the bedside.

This study has several limitations. First, a gold standard for sepsis does not exist. This poses a significant challenge when trying to characterize the definition performance. Instead of comparisons to a gold standard, we are left with comparisons to subjective and objective metrics to characterize the performance of ESP, including physician case review, billing codes, and the outcomes of in-hospital mortality and ICU LOS greater than or equal to 3 days. Each of these metrics has limitations. Physician case review is subjective, with known variation between clinicians in what constitutes sepsis. Although we only report the results of physician review for cases meeting the ESP criteria (the focus of this work), prior assessments of SEP-1 and Sepsis-3 have demonstrated that automation of these definitions results in the identification of many false-positive cases. Our assessments of sensitivity and PPV are based on billing codes and are therefore dependent in part on local billing practices, as well as any variation in sepsis diagnostic practices. Despite this limitation, and the known operational differences around the management of sepsis between the three hospitals contributing data to this study, the performance of each sepsis definition compared with billing codes was similar across all hospitals (Supplemental Table 5, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>). Although it is not clear that the observed BCPV would be reproduced at another site with different practices, we would expect the trends between the different definitions to be similar. Second, although we included patients presenting to the emergency department who were never admitted to the hospital in our cohort, we have little to no knowledge of their outcome after they were discharged. As such, the performance of ESP in this subpopulation is less clear. These cases, however, represent a small percent of the sepsis population (Supplemental Table 3, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>) and are unlikely to significantly impact the results. Finally, although ESP was validated in a diverse patient population using patient records from three hospitals including both academic and community medical centers, all three were part of the same medical system. Although it is unknown how the performance would generalize to other populations, BCS and BCPV were consistent across all hospitals (Supplemental Table 5, Supplemental Digital Content 1, <http://links.lww.com/CCX/A106>). Additional conditions may be needed to adapt the definition to other patient populations; however, this would likely improve rather than erode performance.

CONCLUSIONS

Overall, ASE and ESP identified patients with higher rates of in-hospital mortality and ICU LOS greater than or equal to 3 days and lower rates of comorbidities. By accounting for comorbidities that mimic the signs of sepsis, ESP achieved a higher sensitivity than ASE while maintaining comparable precision. Additional work is needed to automate and expand the list of potentially confounding comorbidities and to further improve the quality of automatic sepsis identification. Finally, although the validations presented here focus on sepsis, ESP could likely be applied to other syndromes without a gold-standard laboratory test.

REFERENCES

1. Bates DW, Saria S, Ohno-Machado L, et al: Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 2014; 33:1123–1131
2. Rajkomar A, Dean J, Kohane I: Machine learning in medicine. *N Engl J Med* 2019; 380:1347–1358
3. Topol EJ: High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25:44–56
4. Bhattacharjee P, Edelson DP, Churpek MM: Identifying patients with sepsis on the hospital wards. *Chest* 2017; 151:898–907
5. Henry KE, Hager DN, Pronovost PJ, et al: A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7:299ra122
6. Futoma J, Hariharan S, Heller K: Learning to Detect Sepsis With a Multitask Gaussian Process RNN Classifier. 2017. Available at: <http://arxiv.org/abs/1706.04152>. Accessed September 27, 2019
7. Umscheid CA, Betesh J, VanZandbergen C, et al: Development, implementation, and impact of an automated early warning and response system for sepsis. *J Hosp Med* 2015; 10:26–31
8. Mao Q, Jay M, Hoffman JL, et al: Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open* 2018; 8:e017833
9. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; 315:801–810
10. Healthcare Cost and Utilization Project (HCUP): *HCUP facts and figures: Statistics on hospital-based care in the United States, 2009*. Rockville, MD, Agency for Healthcare Research and Quality (US), 2011. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22514803>. Accessed September 27, 2019
11. 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; 29:1303–1310
12. Paoli CJ, Reynolds MA, Sinha M, et al: Epidemiology and costs of sepsis in the United States—an analysis based on timing of diagnosis and severity level. *Crit Care Med* 2018; 46:1889–1897
13. 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; 318:1241–1249
14. Soleimani H, Hensman J, Saria S: Scalable joint models for reliable uncertainty-aware event prediction. *IEEE Trans Pattern Anal Mach Intell* 2018; 40:1948–1963
15. Ginestra JC, Giannini HM, Schweickert WD, et al: Clinician perception of a machine learning–based early warning system designed to predict severe sepsis and septic shock. *Crit Care Med* 2019; 47:1477–1484
16. Ruppel H, Liu V: To catch a killer: Electronic sepsis alert tools reaching a fever pitch? *BMJ Qual Saf* 2019; 28:693–696
17. Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; 315:762–774
18. Angus DC, Seymour CW, Coopersmith CM, et al: A framework for the development and interpretation of different sepsis definitions and clinical criteria. *Crit Care Med* 2016; 44:e113–e121
19. Rhee C, Dantes RB, Epstein L, et al: Using objective clinical data to track progress on preventing and treating sepsis: CDC'S new 'adult sepsis event' surveillance strategy. *BMJ Qual Saf* 2019; 28:305–309
20. Rhee C, Jentzsch MS, Kadri SS, et al: 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
21. Rhee C, Gohil S, Klompas M: Regulatory mandates for sepsis care—reasons for caution. *N Engl J Med* 2014; 370:1673–1676.
22. Levy MM, Fink MP, Marshall JC, et al: 2001 SCCM/ESICM/ACCP/ATLS/SIS International sepsis definitions conference. *Intensive Care Med* 2003; 29:530–538
23. Rhee C, Zhang Z, Kadri SS, et al: CDC Prevention Epicenters Program. Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 Sequential Organ Failure Assessment criteria. *Crit Care Med* 2019; 47:307–314
24. Rhee C, Filbin MR, Massaro AF, et al: Compliance with the national SEP-1 quality measure and association with sepsis outcomes. *Crit Care Med* 2018; 46:1585–1591
25. Lewis D: Causation. *J Philos* 1973; 73:556–567
26. Makam AN, Nguyen OK, Auerbach AD: Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review. *J Hosp Med* 2015; 10:396–402
27. Alsolamy S, Al Salamah M, Al Thagafi M, et al: Diagnostic accuracy of a screening electronic alert tool for severe sepsis and septic shock in the emergency department. *BMC Med Inform Decis Mak* 2014; 14:105.
28. Seymour CW, Coopersmith CM, Deutschman CS, et al: Application of a framework to assess the usefulness of alternative sepsis criteria. *Crit Care Med* 2016; 44:e122–e130