Leveraging electronic health record data to improve sepsis surveillance

Claire N Shappell, Chanu Rhee

Sepsis, the syndrome of life-threatening organ dysfunction that complicates severe infection, is a leading cause of death and disability worldwide. A growing recognition of the enormous burden of sepsis has spurred numerous awareness campaigns, quality improvement initiatives and regulatory measures in recent years. Reliably tracking the burden of sepsis is challenging, however, because sepsis is a clinical syndrome based on a constellation of non-specific signs and symptoms and lacks a gold standard for diagnosis. Given the substantial resources being dedicated to improving sepsis care and outcomes, a parallel investment in developing robust, high-quality surveillance tools is necessary to understand which initiatives are effective and where best to allocate future resources.

Until recently, sepsis surveillance has primarily been conducted using hospital discharge diagnosis codes. Epidemiological studies using these data have consistently shown dramatic increases in sepsis incidence and declines in case fatality rates over the past several decades. However, this method is seriously flawed since it requires (1) clinicians to recognize sepsis by identifying that infection is present and responsible for organ dysfunction; (2) clinicians to document sepsis in the medical record and (3) hospital coders to appropriately identify this documentation and assign sepsis as a primary or secondary diagnosis. These steps are subjective and easily biased by changing diagnosis and coding practices over time. Specifically, education and awareness campaigns, new screening protocols and international guidelines are all constantly encouraging early detection of sepsis and organ dysfunction. This, by design, leads to the diagnosis of ‘sepsis’ in more mildly ill patients that previously might only have been labelled by their specific infection (eg, pneumonia) or non-specific illnesses. In the USA, where sepsis diagnoses are tied to the highest level of patient complexity and reimbursement, hospitals also have a clear financial incentive to code for sepsis. Diagnosing earlier and milder forms of sepsis may benefit patients, but it creates an ascertainment bias for surveillance since it is difficult to know whether the reported increases in sepsis incidence and declining mortality rates reflect true changes in disease epidemiology and better sepsis care, or simply artefacts from the inclusion of more patients with less severe illness in the denominator.

Some healthcare systems have used prospective registries based on various screening protocols to track sepsis outcomes. However, this method is also vulnerable to ascertainment bias since the implementation of these screens tends to enhance early identification of sepsis and therefore also captures increasingly milder forms of sepsis. Prospective registries are also resource-intensive and have limited comparability across hospitals and geographical regions due to heterogeneous inclusion criteria. Death records are another data source that have been used to generate national and global estimates of sepsis mortality, but physicians are notoriously inaccurate at coding causes of death and sepsis in particular tends to be under-coded. Furthermore, trends in the coding of sepsis on death certificates are subject to the same changes in diagnosis and documentation practices as hospital administrative data.

The need for a more objective, consistent and scalable approach to sepsis surveillance has recently led some researchers and policymakers to turn to direct clinical indicators of sepsis that can be extracted from electronic health record (EHR) systems which are increasingly ubiquitous in the USA and other developed countries. A prominent example of this
approach is the ‘Adult Sepsis Event’ (ASE) definition created by the US Centers for Disease Control and Prevention in 2018. The ASE was conceptually based on the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) framework of sepsis as infection with concurrent organ dysfunction, but was optimised for retrospective surveillance across a broad range of hospitals using data routinely available in EHRs rather than for real-time decision-making. The ASE identifies hospitalisations with presumed serious infection, as defined by a blood culture order and administration of at least 4 days or antibiotics (or fewer in cases of death, discharge to hospice, transfer to another acute care hospital, or transition to comfort measures before 4 days), and concurrent acute organ dysfunction, defined as initiation of vasopressors or mechanical ventilation, elevated lactate, or clearly defined changes in creatinine, total bilirubin or platelets from patients’ baseline values. ASE requires four antibiotic days to improve specificity by excluding patients who only briefly receive empiric antibiotics and also mitigate potential bias from increased screening and decreasing thresholds to start empiric antibiotics for suspected sepsis. The ASE organ dysfunction criteria resemble the Sequential Organ Failure Assessment (SOFA) score used by Sepsis-3, but use binary thresholds and a smaller number of data elements for greater simplicity to enable use in a wide range of hospitals and EHR systems (table 1).

The ASE definition was initially developed as part of a 2017 multicenter study of the burden of sepsis in the USA and applied across a nationally representative cohort of 409 diverse hospitals from seven datasets. This study yielded a sepsis prevalence rate of 6% in hospitalised adult patients and an in-hospital mortality rate of 15%; when extrapolated nationwide, this generated an estimated 1.7 million adult sepsis cases and 270,000 associated deaths. On medical record reviews, ASE criteria had reasonable sensitivity (69.7%) and good specificity (98.1%) compared with the clinical Sepsis-3 definition. Many of the false positives and false negatives, however, were due to intentional mismatches between the ASE organ dysfunction criteria and the SOFA score used by the Sepsis-3 definition, as the ASE criteria were designed to simplify the number of data elements to facilitate consistent implementation across different EHR systems (eg, by identifying respiratory failure by mechanical ventilation alone rather than PaO2/FiO2 ratios, using any vasopressor initiation rather than specific vasopressor doses, and excluding Glasgow Coma Scale scores). Therefore, the ‘accuracy’ of ASE depends on whether one truly considers Sepsis-3 to be the ‘gold standard’ for sepsis diagnosis. When used to examine sepsis incidence and mortality from 2009 to 2014, the ASE definition generated much more stable trends compared with administrative definitions, and in fact no significant change in incidence or combined death or discharge to hospice was seen when the lactate criteria was omitted (an a priori decision due to increased lactate testing over the period studied).

The ASE was the beginning of an important paradigm shift towards population-level sepsis surveillance using EHR data, but it is certainly not the end. In this issue of BMJ Quality & Safety, Valik and colleagues present the first validation of an EHR-based algorithm based directly on Sepsis-3 criteria and its application to measure sepsis incidence, mortality and variation across non-ICU wards in a Swedish academic medical centre. As per the work by Seymour and the Sepsis-3 task force, suspected infection was defined as any culture obtained (not just blood cultures) and at least two doses of antimicrobials administered, while organ dysfunction was defined by a rise in maximum SOFA score around the time of infection onset by at least two points compared with a baseline SOFA score (table 1). On medical record review, this algorithm achieved very high sensitivity (88.7%), specificity (98.5%) and positive predictive value (PPV) (88.1%) relative to Sepsis-3 criteria as determined by two infectious disease physicians. The performance was excellent across both community-onset and hospital-onset sepsis—an important finding given that administrative data can only distinguish these two conditions by present-on-admission codes, which are often inaccurate and variably applied across hospitals. Sensitivity analyses using alternative definitions of suspected infection, including blood cultures and 4 days of antibiotics as in the ASE definition, had lower sensitivity (71.8% for the ASE equivalent) though improved specificity and PPV (99.2% and 91.7%, respectively). When the algorithm was applied to the hospital’s population over a 1.5-year period, it identified 10.4% of patients as septic (1.3% hospital-onset and 9.1% community-onset sepsis), with an in-hospital mortality rate of 8.6%.

This study provides further evidence that EHR data can be used to build an accurate automated sepsis surveillance system, and is the first medical record-based validation of an algorithm based directly on the SOFA score and Sepsis-3 criteria. The mortality rate of 8.6% is substantially lower than ASE’s mortality rate but is close to the 10% rate in US cohorts used for the derivation and validation of Sepsis-3 criteria, suggesting at least some degree of generalisability. As the authors assert, the Sepsis-3 algorithm identifies a less severely ill set of patients than ASE and therefore may be more relevant for surveillance of general (non-ICU) wards. Furthermore, while the requirement for only two doses of antimicrobials in their definition of suspected infection may cost some specificity, it allows for the possibility of prospective monitoring of sepsis cases as they develop in the hospital and influencing real-time clinical decision-making to improve sepsis care.
Despite these promising results, there are some caveats to this study that are worth noting. First, the algorithm was studied and validated in a single-centre population with a much lower burden of comorbid conditions compared with the multi-centre cohort in which the ASE was studied; it is therefore unclear whether the Sepsis-3 algorithm would maintain its high specificity if applied to a more medically complex population with a greater prevalence of pre-existing organ dysfunction. Second, ICU time was censored due to a lack of data on vital signs and medications, and so their estimations of sepsis incidence should be interpreted with caution. Third, the extent to which the algorithm is susceptible to ascertainment bias from changing clinical practice over time (and changing data availability in EHRs) is unknown since the authors did not use it to track sepsis trends in their hospital.

More broadly, it is important to consider where the automated Sepsis-3 algorithm fits in the framework of sepsis definitions. Given the complexity of sepsis, it is susceptible to ascertainment bias from changing clinical practice over time (and changing data availability in EHRs) is unknown since the authors did not use it to track sepsis trends in their hospital.

### Table 1: Comparison of automated Sepsis-3 algorithm and CDC adult sepsis event criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Sepsis-3 algorithm as implemented by Valik et al</th>
<th>CDC Adult Sepsis Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infection</td>
<td>1. Any clinical culture obtained, AND 2. ≥2 antibiotic doses within 6–48 hours (Culture sites include abdomen, blood, bone, bronchoalveolar lavage, cerebral spinal fluid, catheters/devices, nasopharynx, pleural space, skin/tissue, sputum, stool, synovial fluid, urine. Cultures types include bacterial or C. difficile toxin, Mycoplasma pneumoniae DNA, enterohemorrhagic E. coli DNA, Legionella urine antigen, fungal cultures from blood. If antibiotic administration occurred first, a culture must be obtained within 24 hours. If a culture was obtained first, an antibiotic must be given within 72 hours. One antibiotic dose permitted if patient was admitted to the ICU prior to 24 hours, or died prior to 48 hours from the first antibiotic dose. 'Onset of infection' defined as the first of either event)</td>
<td>1. Blood culture obtained, AND 2. ≥4 consecutive antibiotic days</td>
</tr>
<tr>
<td>Organ Dysfunction</td>
<td>Increase in modified SOFA Score by ≥2 points from baseline during window of up to 48 hours before to 24 hours after onset of infection:</td>
<td>≥1 of the following &quot;eSOFA&quot; criteria within +/-2 calendar days of blood culture day:</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>1. Mean arterial pressure &lt;70 mmHg (Baseline=last measured mean arterial pressure before suspected infection window (only during current hospitalisation). Vasopressor doses not used since surveillance performed outside the ICU)</td>
<td>Vasopressor initiation</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>1. PaO2/FiO2&lt;400 or SpO2/FiO2&lt;512 2. PaO2/FiO2&lt;300 or SpO2/FiO2&lt;257 3. PaO2/FiO2&lt;200, or SpO2/FiO2&lt;214 4. PaO2/FiO2&lt;100, or SpO2/FiO2&lt;89 (Baseline=last PaO2 or SpO2 prior to suspected infection window during last 3 months. ICD codes for home oxygen or ventilator use in prior year=2 baseline points)</td>
<td>Mechanical ventilation initiation</td>
</tr>
<tr>
<td>Renal</td>
<td>1. Creatinine 110–170 µmol/L 2. Creatinine 171–299 µmol/L 3. Creatinine 300–440 µmol/L 4. Creatinine≥440 µmol/L (Baseline=last measured creatinine prior to suspected infection window during last 3 months. ICD codes for chronic dialysis=4 baseline points. Urine output not used due to data availability)</td>
<td>≥2x Creatinine or ≥50% of estimated glomerular filtration rate relative to baseline (Baseline creatinine=lowest during hospitalisation if infection onset on hospital day≤2, or lowest during ≥2-day window period if infection onset on hospital day&gt;2. Patients with ICD codes for end-stage renal disease excluded)</td>
</tr>
<tr>
<td>Hepatic</td>
<td>1. Bilirubin 20–32 µmol/L 2. Bilirubin 33–101 µmol/L 3. Bilirubin 102–204 µmol/L 4. Bilirubin≥204 µmol/L (Baseline=last measured bilirubin prior to suspected infection window during last 3 months)</td>
<td>Bilirubin≥2.0 mg/dL and ≥2x from baseline (Baseline bilirubin=lowest during hospitalisation if infection onset on hospital day≤2, or lowest during ≥2-day window period if infection onset on hospital day&gt;2)</td>
</tr>
<tr>
<td>Coagulation</td>
<td>1. Platelets 100–149×10^3/µL 2. Platelets 50–99×10^3/µL 3. Platelets 20–49×10^3/µL 4. Platelets&lt;20×10^3/µL (Baseline=last measured platelet count prior to suspected infection window during last 3 months)</td>
<td>Platelet count&lt;100×10^3/µL and ≥50% decline from baseline (baseline must ≥100) (Baseline platelets=lowest during hospitalisation if infection onset on hospital day≤2, or lowest during ≥2-day window period if infection onset on hospital day&gt;2)</td>
</tr>
<tr>
<td>Neurologic (SOFA) or Perfusion (eSOFA)</td>
<td>1. Glasgow Coma Scale score 13–14 2. Glasgow Coma Scale score 10–12 3. Glasgow Coma Scale score 6–9 4. Glasgow Coma Scale score&lt;6 (Baseline=last measured value before suspected window (only during current hospitalisation). If Glasgow Coma Scale unavailable, structured data on ‘alert’ (0 points) or ‘not alert’ (one point) used)</td>
<td>Lactate≥2.0 mmol/L</td>
</tr>
</tbody>
</table>

CDC, Centers for Disease Control and Prevention; FiO2, fraction of inspired oxygen; PaO2, arterial PaO2 of oxygen; SOFA, Sequential Organ Failure Assessment; SpO2, peripheral capillary oxygen saturation.
no one set of criteria can suit the needs of all stakeholders. For example, clinicians require a definition optimised for sensitivity and ease of application at the bedside to facilitate timely treatment and avoid missing cases. In contrast, a surveillance definition is meant to reliably track sepsis over time and across different settings to characterise changes in disease epidemiology, interpret the impact of prevention and treatment initiatives, benchmark incidence and outcomes across facilities and geographical regions (and thus identify opportunities for improvement), and guide resource and research investments. As such, surveillance definitions typically prioritise specificity, objectivity and reproducibility over timely diagnosis. This sometimes means that ambiguous or mild cases are excluded. Furthermore, a low burden of measurement is important to facilitate widespread implementation.

With these considerations in mind, the automated Sepsis-3 algorithm appears to be very well-suited to track sepsis incidence and outcomes within the hospital where it was developed. However, it is unclear the degree to which consistent implementation of this approach across a diverse range of hospitals is feasible given the relative complexity of Sepsis-3 criteria and wide variability in the sophistication of EHR systems and data repositories. Prior work has demonstrated how seemingly minor variations in the definition and measurement of the traditional systemic inflammatory response syndrome-based sepsis criteria can have a major impact on the apparent incidence of sepsis. For the Sepsis-3 algorithm, identifying all potential clinical cultures as opposed to blood cultures alone (as per ASE criteria) dramatically expands the number of data elements that need to be identified and mapped. Furthermore, the SOFA score is highly sensitive to missing data and includes several elements that are inconsistently measured across hospitals and variably stored in EHRs, such as Glasgow Coma Scores, vasopressor doses, urine output and blood gas data. Indeed, even in this study, missing SOFA score data elements were common (particularly Glasgow Coma Scale and bilirubin), and several modifications of the SOFA score were needed based on data availability, such as use of peripheral capillary oxygen saturation instead of the PaO2 of oxygen and the omission of urine output. While these are relatively minor adaptations, they underscore the likelihood that slight variations in SOFA implementation are likely to occur across hospitals based on data availability, each of which could confound attempts at comparing sepsis rates and outcomes across facilities and geographical regions and measuring the national or international burden of sepsis. This is an important distinction from ASE, which was designed with particular attention to simplicity and ease of adoption across a broad range of hospitals.

Ultimately, the study by Valik and colleagues represents another important step forward in sepsis surveillance as we move further away from reliance on administrative data and towards a more objective approach using clinical data from EHRs to more reliably study changes in epidemiology and better care for sepsis patients. Further validation and comparisons of this Sepsis-3-based algorithm with ASE and other EHR-based definitions across diverse populations and EHR systems are needed to enable hospitals, policymakers and researchers to decide how best to track sepsis incidence and outcomes and tailor surveillance approaches to their particular needs.

**Funding** CR received support from the Agency for Healthcare Research and Quality (K08HS025008).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; internally peer reviewed.

**REFERENCES**

editorial


