Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: observational study using electronic health records data


ABSTRACT

Background Surveillance of sepsis incidence is important for directing resources and evaluating quality-of-care interventions. The aim was to develop and validate a fully-automated Sepsis-3 based surveillance system in non-intensive care wards using electronic health record (EHR) data, and demonstrate utility by determining the burden of hospital-onset sepsis and variations between wards.

Methods A rule-based algorithm was developed using EHR data from a cohort of all adult patients admitted at an academic centre between July 2012 and December 2013. Time in intensive care units was censored. To validate algorithm performance, a stratified random sample of 1000 hospital admissions (674 with and 326 without suspected infection) was classified according to the Sepsis-3 clinical criteria (suspected infection defined as having any culture taken and at least two doses of antimicrobials administered, and an increase in Sequential Organ Failure Assessment (SOFA) score by ≥2 points) and the likelihood of infection by physician medical record review.

Results In total 82,653 hospital admissions were included. The Sepsis-3 clinical criteria determined by physician review were met in 343 of 1000 episodes. Among them, 313 (91%) had possible, probable or definite infection. Based on this reference, the algorithm achieved sensitivity 0.887 (95% CI: 0.799 to 0.964), specificity 0.985 (95% CI: 0.978 to 0.991), positive predictive value 0.881 (95% CI: 0.833 to 0.926) and negative predictive value 0.986 (95% CI: 0.973 to 0.996). When applied to the total cohort taking into account the sampling proportions of those with and without suspected infection, the algorithm identified 8599 (10.4%) sepsis episodes. The burden of hospital-onset sepsis (>48 hour after admission) and related in-hospital mortality varied between wards.

Conclusions A fully-automated Sepsis-3 based surveillance algorithm using EHR data performed well compared with physician medical record review in non-intensive care wards, and exposed variations in hospital-onset sepsis incidence between wards.

INTRODUCTION

Sepsis, a severe organ dysfunction induced by infection, is a leading cause of morbidity and death worldwide. The true burden of sepsis has been difficult to assess, mainly due to the absence of a generalisable gold standard. About one third of sepsis episodes are considered healthcare-associated and the problem of sepsis needs to be addressed as a patient safety concern.

Surveillance with feedback to healthcare personnel and policy makers is the backbone of most quality improvement programmes for healthcare-associated infections. To be useful, such surveillance systems require standardised case-definitions free from subjective interpretations and appropriate denominator data. Surveillance systems based on clinical data are preferred to administrative data, as these are more objective, reproducible and stable over time. Fully-automated surveillance systems with data from electronic health records (EHR) could replace surveillance relying on manual chart review and generate continuous data from large populations, but needs thorough validation before
implementation.5 As mandatory reporting of sepsis is becoming increasingly common, defining a surveillance method that produces objective high quality data is important.9 Reliable sepsis surveillance data can benefit large patient groups by allowing clinical resources to be directed to where they are most needed. Continuous incidence monitoring can also be used to evaluate quality of care interventions down to the ward level, and for benchmarking sepsis prediction models and artificial intelligence tools integrated with the EHR.8 10 11

In 2016 the Third International Consensus Definition for Sepsis and Septic Shock (Sepsis-3) was introduced.12 13 An advantage of the new definition (here denoted Sepsis-3 clinical criteria) lies in its objective case-definition, which requires that the patient has a suspected infection in combination with a newly developed organ dysfunction.13 Yet, the performance of an automated surveillance system using these criteria has not been evaluated. In 2018, the Center of Disease Control and Prevention (CDC) presented an additional sepsis definition called Adult Sepsis Event (ASE) aimed specifically for surveillance purposes.14 The ASE differs from the Sepsis-3 clinical criteria with regards to both classifications of suspected infection and organ dysfunction. The focus of ASE is on patients with more severe disease, with more than half of cases admitted to the intensive care unit (ICU), and it has been shown to underestimate the burden of sepsis cases defined by Sepsis-3 criteria.15

The primary aim of this study was to develop and validate a fully-automated EHR-based surveillance algorithm against physician medical record review in non-intensive care wards using the Sepsis-3 clinical criteria. A secondary aim was to demonstrate the algorithm’s utility by determining the burden of hospital-onset sepsis in a general hospital population.

METHODS

Design, data source and study population
This was an observational study performed at an academic centre with 1350 beds divided between two hospitals and serving a population of 2.3 million inhabitants. Data was obtained from routinely prospectively entered information in the EHR system, stored in a research databank called Health Bank—Swedish Health Record Research Infrastructure.16 The database structure is a duplicate of the operating EHR system, where each subject can be followed over time, and consists of all medical records from more than 2 million anonymised patients that received care at the hospital between 2006 and 2013. Due to improved recording in the EHR during the later years, analyses were restricted to July 2012 until December 2013, except for information about International Classification of Diseases (ICD) codes that were used to estimate the presence of co-morbidities in patients, which were retrieved up to 5 years before inclusion. Data collection included demographics, hospital administrative data, vital parameters, laboratory findings, microbiological data, medications and in-hospital mortality.

Patients >18 years admitted to the hospital for >24 hours were included, and followed until first sepsis episode, discharge or death. Patients were excluded if admitted to an obstetric ward and censored during ICU-care, due to lack of data on vital parameters and medication for these wards.

Sepsis-3 surveillance case definition
The rule-based algorithm was based on the operational Sepsis-3 clinical criteria: a suspected infection in combination with an increase in Sequential Organ Failure Assessment (SOFA) score by ≥2 points compared with the baseline.13 Suspected infection was defined as having any culture taken and at least two doses of antimicrobials administered. If the patient was admitted to the ICU prior to 24 hours, or died prior to 48 hours from the first dose of antimicrobials, they were deemed to have a suspected infection despite only being given one dose. Cultures had to be performed within 24 hours after the start of antimicrobial treatment. Antimicrobial treatment had to be started within 72 hours after culture. Onset of infection was determined based on which of these events occurred first.13 Sensitivity analyses were done using different definitions of suspected infections: only blood cultures and two doses of antimicrobials, any culture and four calendar days of antimicrobials or only blood cultures and four calendar days of antimicrobials, of which the last being equivalent to the ASE definition (online supplementary methods 1).14

Organ dysfunction was measured as the maximum SOFA score 48 hours before to 24 hours after onset of infection and compared with a baseline SOFA score measured separately (online supplementary methods 1 and online supplementary figure 1). Similar to the study that developed the Sepsis-3 clinical criteria, missing values during the 72 hours window were considered to be normal.13 Since we studied a non-ICU population, some modification to the SOFA score was done. The most important changes were, (i) if PaO2 was not available it was calculated from peripheral capillary oxygen saturation (SpO2), (ii) if Glasgow Coma Scale (GCS) was not available, structured data on ‘alert’ (interpreted as GCS score 15 points) or ‘not alert’ (interpreted as GCS score 14 points) was used, and (iii) urine output was not used due to data being unavailable. For each component of the SOFA score, the baseline was defined as the latest value measured before the 72 hours time window, and was assumed to be zero in patients not known to have a pre-existing organ dysfunction. Pre-existing organ dysfunction was based on measured parameters (coagulation, liver, renal and respiration) within the previous three months or a specific ICD-code (chronic dialysis or home oxygen...
therapy) within the last year. For SOFA cardiovascular and central nervous system (CNS) scores, only values measured during the current hospital episode was used. Onset of sepsis was when the patient fulfilled the organ dysfunction criteria.

To capture the burden of healthcare-associated sepsis, hospital-onset (HO) sepsis was defined as onset of suspected infection and organ dysfunction 48 hours after admission, or re-admission with sepsis within 48 hours of discharge. All other episodes were defined as community-onset (CO) sepsis. A patient could have several suspected infections, but only the first episode of sepsis was considered for each hospital episode. Algorithm classification based on clinical data was compared with classification using the following validated criteria based on CDC and The International Sepsis Forum definitions.\(^{17-20}\) Accordingly, one minor modification to the criteria was done in this study. We added unknown source, defined as patients (i) with symptoms of an infection, (ii) the symptoms indicated an infection according to the attending physician, and (iii) the patient received a full course of anti-infective treatment, but (iv) no source could be determined. Unknown source could only be classified as possible infection. In the assessment of the sensitivity and specificity of the surveillance algorithm, patients had to fulfil both the Sepsis-3 clinical criteria and the possible, probable or definite infection criteria to be classified as true sepsis.

**Statistical analyses**

To assess algorithm performance in the intended target population of all patients admitted to the hospital, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated by generalising the proportions from validation to the entire cohort, as previously described by Rhee et al.\(^{15}\) CI for sensitivity, specificity, PPV and NPV were calculated as the 2.5\(^{th}\) and 97.5\(^{th}\) percentiles of point estimates obtained from 10,000 bootstrap samples for each of the two validation sets (n=674 and n=326) using the ‘boot’ package of R. To account for uncertainty, the bootstrapping was performed before extrapolating the proportions from validation to the entire hospital cohort. The extrapolation accounted for the nested selection of suspected CO and HO infections, as well as for the proportion of sepsis cases from the population of patients who did not have a suspected infection episode. In sensitivity analyses of different definitions of suspected infection, the proportion of patients that were falsely categorised as true negatives due to not fulfilling the tested suspected infection definition was also accounted for. When assessing algorithm performance for CO infections, HO infections were omitted and vice versa.

CIs for incidence densities of CO sepsis per 100 admissions and HO sepsis per 1000 patient days at risk were calculated using the R package ‘compere’, assuming a log-normal distribution. Cumulative incidence function (CIF) for the probability of HO sepsis was calculated using the Aalen-Johansen estimator and taking into account competing risks: ICU admission, discharge or death.\(^{21}\) Pairwise comparison of CIF between wards was calculated using the R package ‘cmprsk’, which is based on the Fine and Gray (1999) formula.\(^{22}\) Cumulative incidence regression analysis for in-hospital death stratified by likelihood of infection (none-possible vs probable-definite) was determined with discharge as a competing event using the ‘stcrreg’ command in STATA. Confounders associated with both the likelihood of infection and in-hospital death were considered using a directed acyclic graph based on a priori clinical expertise and available literature. Accordingly, adjustments were made for age, Charlson comorbidity index and CO/HO onset, but not for severity of disease since this covariate could be on the causal pathway between likelihood of infection and in-hospital death.
Results are presented as median (med) and IQR or numbers (No.) and percentages as appropriate. Differences between categorical variables and between continuous variables were assessed using the Fisher exact and Mann-Whitney U tests, respectively. Missing data was not considered missing at random, but rather owing to clinical decisions. Based on the assumption that data would still be not missing at random even after controlling for other observed variables, we did not perform multiple imputation for individual SOFA score variables. Data handling and automated sepsis annotation was performed in Python V.3.6. Statistical analyses were done in R V.3.4.3 and STATA V.14.2.

RESULTS
In total 144 179 hospital admissions of 99 864 patients were recorded during the study period, of which 95 858 admissions fulfilled the inclusion criteria (patients >18 years admitted to the hospital for ≥24 hours). Twelve thousand nine hundred-fifty admissions to the obstetrical wards, 214 admissions to paediatric wards and 41 direct transfers between hospital ICUs were excluded. Finally, a total of 82 633 hospital admissions of 54 884 patients were included in the analysis. There was a suspected infection in 19 479 (23.6%) of the admissions and no suspected infection in 63 174 (76.4%) of the admissions. The median patient age was 64 years, 50.9% were women and the median length of stay was 3.8 days.

Validation of the surveillance case definition
In total, 340 of 674 patients with suspected infection (50.4%) and three of 326 without suspected infection (0.9%) fulfilled Sepsis-3 criteria according to physician medical record review (table 1). Among subjects that fulfilled Sepsis-3 criteria, 109/343 (31.8%) fulfilled the criteria for possible infection, 87/343 (25.4%) for probable infection, and 117/343 (34.1%) for definite infection. In total, 30/343 (8.7%) were classified as no infection. Hence, 311 of 674 patients (46.1%) with suspected infection and two of 326 patients (0.6%) without suspected infection were finally considered as true sepsis by reviewers. In subjects with suspected infection, the algorithm classified 288 true positive, 39 false positive, 324 true negative and 23 false negative sepsis cases. In subjects without suspected infection, the algorithm classified zero sepsis cases resulting in 324 true negative and two false negative cases. Based on the medical record reviewed reference, the algorithm achieved sensitivity 0.887 (95% CI: 0.799 to 0.964), specificity 0.985 (95% CI: 0.978 to 0.991), PPV 0.881 (95% CI: 0.833 to 0.926) and NPV 0.986 (95% CI: 0.973 to 0.996) when extrapolating proportions to the entire hospital cohort (table 2). When assessed only in subjects with suspected infection, the algorithm achieved sensitivity 0.926 (95% CI: 0.896 to 0.953), specificity 0.893 (95% CI: 0.859 to 0.923), PPV 0.881 (95% CI: 0.833 to 0.926) and NPV 0.934 (95% CI: 0.895 to 0.969) (table 2). The most common reasons for misclassification resulting in reduced sensitivity was respiratory or CNS dysfunction only being mentioned in free text, followed by overestimation of pre-existing organ dysfunction or development of infection related organ dysfunction outside of the 72 hours time window (online supplementary table 1). Reasons for imperfect specificity was due to episodes judged by reviewers as no infection, misclassification of baseline SOFA score or obvious measurement errors of vital parameters in the EHR.

For CO-sepsis the algorithm achieved sensitivity 0.910 (95% CI: 0.825 to 0.984), specificity 0.987 (95% CI: 0.982 to 0.991), PPV 0.881 (95% CI: 0.844 to 0.917) and NPV 0.990 (95% CI: 0.980 to 0.998). For HO-sepsis the algorithm achieved sensitivity 0.794 (95% CI: 0.683 to 0.889), specificity 0.997 (95% CI: 0.995 to 0.999), PPV 0.877 (95% CI: 0.782 to 0.966) and NPV 0.994 (95% CI: 0.991 to 0.997). Restricting analyses to hospital admissions without ICU admission (n=78 318) resulted in slightly decreased sensitivity 0.879 (0.793–0.952) but increased specificity 0.988 (0.983–0.992) and PPV 0.895 (0.860–0.931). For hospital admissions with ICU admission (n=4335), sensitivity was higher 0.952 (0.881–1.000) at the expense of decreased specificity 0.938 (0.907–0.969) and PPV 0.800 (0.712–0.894) (online supplementary table 2).

Classification of infection as probable or definite was not associated with a significantly different in-hospital mortality compared with subjects with no or possible infection (online supplementary figure 2). The most common source of infection in true sepsis patients was respiratory (n=119/313, 38.0%), followed by urogenital (n=54/313, 17.3%), unknown source (n=42/313, 13.4%), bloodstream (35/313, 11.2%), skin, bone and joint (30/313, 9.6%), abdominal (n=26/313, 8.3%) and other infectious sources (7/313, 2.2%) (table 1 and online supplementary figure 3). Among patients classified as having an unknown source of infection, 15/42 (35.7%) had neutropenia.

The burden of sepsis
The surveillance algorithm identified 8599 sepsis episodes (10.4% of all hospital admissions), of which 7493 (87.1%) were CO sepsis and 1106 (12.9%) were HO sepsis (table 3). The most common SOFA score triggers for sepsis were respiratory and renal dysfunction (online supplementary figure 4). Availability of data to calculate SOFA score during suspected infection ranged between 92.0%–95.0% (coagulation, renal, respiratory and cardiovascular) and 38.3%–55.0% (liver and CNS) for community-onset episodes, compared with 73.2%–86.2% (coagulation, respiratory, renal and cardiovascular) and 3.0%–30.2% (CNS and liver) for hospital-onset episodes (online supplementary table 3). Assumptions of normal baseline SOFA score were almost exclusively done in community-onset episodes.
**Table 1** Characteristics of patients fulfilling Sepsis-3 clinical criteria* according to physician review of medical records

<table>
<thead>
<tr>
<th>Characteristics of patients fulfilling Sepsis-3 clinical criteria*</th>
<th>All</th>
<th>None</th>
<th>Possible</th>
<th>Probable</th>
<th>Definite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, No. (% of all)</td>
<td>343 (100.0)</td>
<td>30 (8.7)</td>
<td>109 (31.8)</td>
<td>87 (25.4)</td>
<td>117 (34.1)</td>
</tr>
<tr>
<td>Female sex, No. (%)</td>
<td>158 (46.1)</td>
<td>13 (43.3)</td>
<td>43 (39.4)</td>
<td>34 (39.1)</td>
<td>68 (58.1)</td>
</tr>
<tr>
<td>Age, med (IQR)</td>
<td>71 (60–81)</td>
<td>68 (52–81)</td>
<td>70 (60–79)</td>
<td>72 (60–83)</td>
<td>72 (62–81)</td>
</tr>
<tr>
<td>Length of stay, med (IQR)</td>
<td>11 (6–20)</td>
<td>10 (6–14)</td>
<td>10 (6–19)</td>
<td>11 (5–24)</td>
<td>12 (7–22)</td>
</tr>
<tr>
<td>Charlson Comorbidity Index, med (IQR)</td>
<td>2 (0–3)</td>
<td>1 (0–2)</td>
<td>2 (1–3)</td>
<td>1 (0–2)</td>
<td>2 (0–3)</td>
</tr>
<tr>
<td>Risk factors at sepsis onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior surgery (30 days)†, No. (%)</td>
<td>82 (23.9)</td>
<td>6 (20.0)</td>
<td>32 (29.4)</td>
<td>11 (12.6)</td>
<td>33 (28.2)</td>
</tr>
<tr>
<td>Central venous catheter, No. (%)</td>
<td>87 (25.4)</td>
<td>8 (26.7)</td>
<td>33 (30.3)</td>
<td>13 (14.9)</td>
<td>33 (28.2)</td>
</tr>
<tr>
<td>Urinary catheter, No. (%)</td>
<td>80 (23.4)</td>
<td>11 (36.7)</td>
<td>28 (25.7)</td>
<td>18 (20.7)</td>
<td>23 (19.7)</td>
</tr>
</tbody>
</table>

*Sepsis-3 defined as any culture taken and administration of two doses antimicrobials combined with change in SOFA score of two points or more during 48 hours before and 24 hours after onset of infection compared with baseline SOFA score calculated before this time window.
†Weighted Charlson comorbidity index. Total modified Charlson score: 0–24.
‡Prior surgery was generated using procedure codes. Time zero was sepsis onset.
§Hospital-onset defined as an episode 48 hours after admission or if the patient was readmitted with sepsis within 48 hours of discharge. All other episodes were defined as community-onset.
¶ICU admission at any time during hospitalisation.
**Shock defined as patients receiving vasopressors.
††Neutropenia defined as absolute neutrophil count less than 0.5x10⁹/L.
‡‡Bloodstream infection at any time during hospitalisation (online supplement methods 2).

(range between 20.2%–46.8% for liver, renal, coagulation and respiration), but a large portion of patients had a measured baseline value (range between 18.1%–56.0% for liver, respiration, coagulation and renal) (online supplementary table 4). In hospital-onset suspected infection, only 0.4%–3.2% (all SOFA score components) had an assumed normal baseline value.

Only 13.4% of sepsis episodes had an ICD-10 code indicating sepsis. The in-hospital mortality was 8.6% for all sepsis episodes, 8.0% for CO sepsis and 12.7% for HO-sepsis, compared with 2.4% in the entire hospital cohort. The incidence was 9.1 (95% CI: 8.9 to 9.3) per 100 admissions for CO sepsis and 2.6 (95% CI: 2.4 to 2.8) per 1000 patient days for HO sepsis, with a CIF of 0.013 at day 30 for HO sepsis in the competing risk model (online supplementary figure 5). The cumulative incidence of HO sepsis varied significantly depending on type of hospital ward, with the highest risk in Transplant (CIF=0.078) and Haematology (CIF=0.061) wards, and the lowest risk in Orthopaedic (CIF=0.004) wards (figure 1 and online supplement figures 6 and 7). In-hospital mortality
Table 2  Performance of the surveillance algorithm using different definitions of suspected infection

<table>
<thead>
<tr>
<th>Definition of suspected infection</th>
<th>Entire hospital cohort (n=82 653)</th>
<th>Suspected infection validation cohort (n=674)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (95% CI)</td>
<td>Specificity (95% CI)</td>
</tr>
<tr>
<td>Any culture and two doses of antimicrobials (equivalent to Sepsis-3 clinical criteria)</td>
<td>0.887 (0.799 to 0.964)</td>
<td>0.985 (0.978 to 0.991)</td>
</tr>
<tr>
<td>Any culture and 4 days of antimicrobials</td>
<td>0.853 (0.761 to 0.935)</td>
<td>0.988 (0.981 to 0.993)</td>
</tr>
<tr>
<td>Blood culture and two doses of antimicrobials</td>
<td>0.739 (0.639 to 0.832)</td>
<td>0.990 (0.985 to 0.995)</td>
</tr>
<tr>
<td>Blood cultures and 4 days of antimicrobials (equivalent to Adult Sepsis Event definition)</td>
<td>0.718 (0.615 to 0.813)</td>
<td>0.992 (0.986 to 0.996)</td>
</tr>
</tbody>
</table>

NPV, negative predictive value; PPV, positive predictive value.

after HO sepsis was highest in internal medicine wards (17.3%) compared with only 2.1% in thoracic surgery wards (online supplementary table 5).

Sensitivity analysis demonstrating variations in number of sepsis episodes and in-hospital mortality using different definitions of suspected infection are shown in figure 2. Mortality in sepsis episodes fulfilling only the definition used in the Sepsis-3 clinical criteria (any culture and two doses of antimicrobials), but not the definition used in the ASE (blood cultures and four days of antimicrobials), was 8.4% (n=174/2066). This was not significantly different from mortality 8.6% (n=563/6533) in sepsis episodes fulfilling both definitions (p=0.78 for difference).

DISCUSSION

In this study, we show that it is possible to build a fully-automated sepsis surveillance system based on the Sepsis-3 clinical criteria that correctly captures almost 90% of sepsis episodes occurring outside the ICU and assigns the events to space (ward) and time (onset). The mortality and patient characteristics in our study were similar to the studies used when developing the Sepsis-3 definition, speaking in favour of our results being generalisable to the European and US setting. The usefulness of the algorithm was shown by indicating variations in HO sepsis incidence and mortality depending on ward type, which can be used to inform infection prevention interventions and improve sepsis care.

The sepsis definition is based on the pathophysiological response to an infection and is neither constrained to a certain type of infection nor does it require that the infection is microbologically confirmed. Quality improvement initiatives focusing on education and sepsis care bundles have been associated with survival benefits, warranting structured approaches in sepsis care.

The sepsis definition is based on the pathophysiological response to an infection and is neither constrained to a certain type of infection nor does it require that the infection is microbologically confirmed. Quality improvement initiatives focusing on education and sepsis care bundles have been associated with survival benefits, warranting structured approaches in sepsis care. In our study, the majority of patients presented with sepsis on admission, but the burden of HO sepsis was still substantial. Recent data have associated HO sepsis with mortality approximately twice as high compared with CO sepsis. Despite this, traditional surveillance programmes for healthcare-associated infections, such as CDC/National Healthcare Safety Network and European Centre for Disease Prevention and Control, do not include sepsis as a distinct entity. Quality improvement initiatives focusing on education and sepsis care bundles have been associated with survival benefits, warranting structured approaches in sepsis care.

Initiatives to monitor sepsis incidence have often focussed on using administrative hospital data, such as discharge diagnosis, trigger based audits or reporting to clinical databases, all carrying risk of bias and making comparisons between hospitals difficult. The use of ICD-codes for sepsis surveillance is associated with considerable uncertainty and studies indicate that some of the increased incidence of sepsis during the last decade can be explained by changes in coding practices. Overall, epidemiological surveillance based on explicit sepsis ICD-codes seems to underestimate...
### Table 3  Characteristics of fully-automated sepsis incidence surveillance in a general hospital population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All</th>
<th>No sepsis</th>
<th>Sepsis-3 clinical criteria</th>
<th>Community-onset* Sepsis-3 clinical criteria</th>
<th>Hospital-onset* Sepsis-3 clinical criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admissions, No.</td>
<td>82653</td>
<td>74054</td>
<td>8599</td>
<td>7493</td>
<td>1106</td>
</tr>
<tr>
<td>Patients, No.</td>
<td>54884</td>
<td>51343</td>
<td>7286</td>
<td>6472</td>
<td>1055</td>
</tr>
<tr>
<td>Female sex, No. (%)</td>
<td>27928</td>
<td>26378</td>
<td>3213 (43.9)</td>
<td>2876 (44.2)</td>
<td>430 (41.4)</td>
</tr>
<tr>
<td>Age, med (IQR)</td>
<td>64.0 (47.0–75.0)</td>
<td>63.0 (47.0–74.0)</td>
<td>70.0 (59.0–80.0)</td>
<td>70.0 (60.0–81.0)</td>
<td>67.0 (54.0–76.0)</td>
</tr>
<tr>
<td>Length of stay, med (IQR)</td>
<td>3.8 (2.0–7.6)</td>
<td>3.3 (1.9–6.9)</td>
<td>8.0 (4.2–15.7)</td>
<td>7.0 (4.0–12.6)</td>
<td>23.0 (14.7–36.7)</td>
</tr>
<tr>
<td>Charlson Comorbidity Index†, med (IQR)</td>
<td>0 (0–2)</td>
<td>0 (0–2)</td>
<td>2 (0–3)</td>
<td>2 (0–2)</td>
<td>2 (0–3)</td>
</tr>
<tr>
<td>Comorbidities‡, No. (%)</td>
<td>928 (50.9)</td>
<td>513 (51.4)</td>
<td>3783 (51.4)</td>
<td>343 (51.4)</td>
<td>500 (47.9)</td>
</tr>
<tr>
<td>Chronic pulmonary disease</td>
<td>2730 (43.3)</td>
<td>2103 (41.4)</td>
<td>692 (9.5)</td>
<td>645 (10.0)</td>
<td>76 (7.2)</td>
</tr>
<tr>
<td>Cancer</td>
<td>14036 (25.6)</td>
<td>13094 (25.5)</td>
<td>2601 (35.7)</td>
<td>2240 (34.6)</td>
<td>514 (48.7)</td>
</tr>
<tr>
<td>Cerebral vascular disease</td>
<td>4061 (7.4)</td>
<td>3783 (7.4)</td>
<td>670 (9.2)</td>
<td>580 (8.9)</td>
<td>123 (11.7)</td>
</tr>
<tr>
<td>Chronic heart failure</td>
<td>2887 (5.3)</td>
<td>2643 (5.1)</td>
<td>792 (10.9)</td>
<td>708 (10.9)</td>
<td>113 (10.7)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>2638 (4.8)</td>
<td>2485 (4.8)</td>
<td>403 (5.5)</td>
<td>361 (5.6)</td>
<td>49 (4.6)</td>
</tr>
<tr>
<td>Connective tissue disease</td>
<td>1471 (2.7)</td>
<td>1360 (2.6)</td>
<td>287 (3.9)</td>
<td>259 (4.0)</td>
<td>42 (4.0)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>1681 (3.1)</td>
<td>1541 (3.0)</td>
<td>442 (6.1)</td>
<td>394 (6.1)</td>
<td>65 (6.2)</td>
</tr>
<tr>
<td>HIV infection</td>
<td>54 (0.1)</td>
<td>46 (0.1)</td>
<td>18 (0.2)</td>
<td>17 (0.3)</td>
<td>2 (0.2)</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>2222 (4.0)</td>
<td>1998 (3.9)</td>
<td>644 (8.8)</td>
<td>565 (8.7)</td>
<td>111 (10.5)</td>
</tr>
<tr>
<td>Liver disease</td>
<td>1281 (2.3)</td>
<td>1173 (2.3)</td>
<td>340 (4.7)</td>
<td>294 (4.5)</td>
<td>62 (5.9)</td>
</tr>
<tr>
<td>Prior surgery (30 days)§, No. (%)</td>
<td>12274 (14.9)</td>
<td>10091 (13.6)</td>
<td>2186 (25.4)</td>
<td>1458 (19.5)</td>
<td>728 (65.8)</td>
</tr>
<tr>
<td>Suspected infection¶, No. (%)</td>
<td>19479 (23.6)</td>
<td>10880 (14.7)</td>
<td>8599 (100.0)</td>
<td>7493 (100.0)</td>
<td>1106 (100.0)</td>
</tr>
<tr>
<td>Sepsis**, No. (%)</td>
<td>8599 (10.4)</td>
<td>0 (0.0)</td>
<td>8599 (100.0)</td>
<td>7493 (100.0)</td>
<td>1106 (100.0)</td>
</tr>
<tr>
<td>ICD-10 coded</td>
<td>2055 (2.5)</td>
<td>907 (1.2)</td>
<td>1148 (13.4)</td>
<td>939 (12.5)</td>
<td>209 (18.9)</td>
</tr>
<tr>
<td>Community-onset sepsis, No. (%)</td>
<td>7493 (9.1)</td>
<td>0 (0.0)</td>
<td>7493 (87.1)</td>
<td>7493 (100.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Hospital-onset sepsis, No. (%)</td>
<td>1106 (1.3)</td>
<td>0 (0.0)</td>
<td>1106 (12.9)</td>
<td>0 (0.0)</td>
<td>1106 (100.0)</td>
</tr>
<tr>
<td>ICU admission††, No. (%)</td>
<td>4335 (5.2)</td>
<td>3471 (4.7)</td>
<td>864 (10.0)</td>
<td>578 (7.7)</td>
<td>286 (25.9)</td>
</tr>
<tr>
<td>ICU days, med (IQR)</td>
<td>1.4 (0.9–4.1)</td>
<td>1.2 (0.9–3.5)</td>
<td>2.8 (1.1–6.3)</td>
<td>2.1 (1.0–4.6)</td>
<td>4.8 (2.0–11.4)</td>
</tr>
<tr>
<td>Bloodstream infection‡‡, No. (%)</td>
<td>2659 (3.2)</td>
<td>1104 (1.5)</td>
<td>1555 (18.1)</td>
<td>1279 (17.1)</td>
<td>276 (25.0)</td>
</tr>
<tr>
<td>In-hospital mortality, No. (%)</td>
<td>1953 (2.4)</td>
<td>1216 (1.6)</td>
<td>737 (8.6)</td>
<td>596 (8.0)</td>
<td>141 (12.7)</td>
</tr>
</tbody>
</table>

*Hospital-onset defined as a sepsis episode 48 hours after admission or if the patient was readmitted with sepsis within 48 hours of discharge. All other sepsis episodes were defined as community-onset.
†Weighted Charlson comorbidity index. Total modified Charlson score: 0–24.
‡Comorbidity defined according to International Classification of Diseases (ICD)-10 codes registered within five years prior to hospitalisation.
§Prior surgery was generated using procedure codes. For sepsis patients time zero was sepsis onset. For non-sepsis patients time zero was admission.
¶Defined as any culture taken and administration of two doses antimicrobials.
**Sepsis-3 defined as suspected infection combined with change in SOFA score of two points or more.
††Intensive care unit admission any time during hospitalisation.
‡‡Bloodstream infection any time during hospitalisation (online supplement methods 2).

the incidence of sepsis compared with using clinical data, and in our study only 13.4% of sepsis patients had an ICD-code indicating sepsis. Similar findings have been observed in studies comparing medical record review to ICD-codes, but manual medical record review is both resource intensive and associated with subjectivity and limited inter-rater agreement. Recently, a case definition, ASE, was developed by CDC to facilitate automated sepsis surveillance using clinical data from EHR. Compared with Sepsis-3, the ASE algorithm is based on different criteria for both suspected infection and organ dysfunction and tends to capture a patient population with higher mortality than the Sepsis-3 criteria. The sensitivity and specificity of the ASE definition, when using Sepsis-3 as the reference standard, was 69.7% and 98.1% in a US
Figure 1  Cumulative incidence function (CIF) curves of hospital-onset sepsis stratified by ward type and taking into account competing risks ICU-admission, discharge or death. The CIF curves differed significantly in pairwise comparison (online supplementary figure 6). ICU, intensive care unit.

hospital setting,\textsuperscript{15} compared with 88.7\% and 98.5\% for our algorithm in a European hospital. Using the ASE definition of suspected infection (blood culture and four days of antimicrobials) in our cohort resulted in 71.8\% sensitivity and 99.2\% specificity.

In this study, 91\% of patients with sepsis according to the Sepsis-3 clinical criteria had either a possible, probable or definite infection as determined by physician review of medical records in post-hoc assessment, which is similar to a previous report from the ICU.\textsuperscript{17} This suggests that the Sepsis-3 criteria perform well in capturing a patient population where clinicians maintain a suspicion of infection also after the initial treatment phase. Organ dysfunction in the Sepsis-3 clinical criteria is determined by SOFA score and concerns have been raised that this is not suited for EHR-based surveillance due to the inclusion of parameters not frequently measured in most patients.\textsuperscript{42} However, integration of automated SOFA score calculators in EHR systems have shown strong agreement with manual score calculations,\textsuperscript{43} limiting the need to use other criteria for organ dysfunctions. The SOFA score is based on assessment of six organ systems, compared with ASE that assesses five organ systems (CNS dysfunction is omitted). For respiratory and cardiovascular dysfunction, the ASE requires initiation of mechanical ventilation and vasopressor treatment. This biases sepsis surveillance towards patients eligible for aggressive treatment and access to ICU care, limiting generalisability to all hospitalised patients. One of the arguments for abandoning the Sepsis-3 definition in ASE was to facilitate widespread use to hospitals with limited collection of EHR data. However, the only additional data used in our surveillance case definition was vital parameters, which are routinely collected in many hospitals.

We show that it is feasible to use a surveillance algorithm based on the Sepsis-3 clinical criteria to automatically identify sepsis with high sensitivity and predictive values in non-ICU wards, and thus keeping a uniform sepsis definition for EHR surveillance. The objective of such surveillance is not early bedside sepsis recognition, but rather making continuously collected data on disease burden and patient management easily available. The possible use-cases of such surveillance data are multifaceted. First, incidence data presented down to the single-ward level as shown in this study, creates important feedback loops, which can guide quality improvement interventions, such as education programmes, systems for earlier sepsis recognition, treatment bundles and targeted infection control measures. Second, since Sepsis-3 based surveillance criteria do not require four days of antimicrobial treatment, but two doses, feedback on patients that have developed sepsis can be presented to clinicians early in the treatment course. This facilitates optimisation of care beyond the very initial treatment phase, such as better source control, adequate diagnostics, optimised antimicrobial treatment, infectious diseases specialist consultation and targeted rehabilitation, all of which have the potential to improve patient outcomes.\textsuperscript{44, 45}
addition, 9% of patients fulfilling the Sepsis-3 clinical criteria did not have an infection, and 32% had only a possible infection, indicating the possibility to use this type of surveillance system as part of an antimicrobial stewardship programme, to balance the empirical broad spectrum antimicrobial treatment imposed by guidelines such as the Surviving Sepsis Campaign Bundle.46–48

**Strength and limitations**
A strength of our study is the use of a large clinical dataset representative of the population in a defined catchment area. This is, to our knowledge the first report of a sepsis surveillance system using the Sepsis-3 clinical criteria as case definition, which overlaps better with other standards used for early sepsis recognition. This enables the integration of surveillance data in the direct clinical care of individual patients which can encourage clinicians to use the data, as opposed to implementing criteria developed exclusively for retrospective surveillance and thus risking to disconnect surveillance from the everyday clinical work. When developing and validating the algorithm, we used a duplicate of the EHR system, to ensure that our model can be implemented using real-time patient data. We could follow each subject over time and were not limited to data from the current hospital admission. This improves proper calculation of baseline organ dysfunction, which has been a limitation in previous methods.14 Furthermore, we performed medical record review, showing that a rule-based surveillance algorithm performed well in non-ICU wards where data is usually of lower resolution and quality. This demonstrates that automated sepsis surveillance using the Sepsis-3 clinical criteria can be done without the need for complex computational methods such as text mining of unstructured data in EHR notes.

Limitations of fully-automated surveillance systems include possible misclassification of sepsis since both the algorithm and validation with medical record review depends on correct and accessible data in the EHR system. Not all hospital admissions with suspected infection contained the measurements necessary to assess a complete SOFA score, leading to missing data. By definition in the Sepsis-3 clinical criteria, missing values of SOFA score components were assumed to be normal, which may have affected correct classification of organ dysfunction and sepsis. Even though our validation sample included hospital admissions with and without suspected infection, our reference standard was based on infections recognised by clinicians and we may have missed sepsis cases among patients where an infection passed unnoticed. Sepsis classification can be affected by updates and changes in the EHR system, as well as by differences in recordings and access of data between wards, which could have influenced our results. This may explain the decreased algorithm specificity and PPV when restricting analyses to only hospital admissions including an ICU admission, from where we did not have access to data on medications and vital parameters. Since we did not include patient risk-time while in ICUs or obstetrical wards, our results cannot be generalised to such settings and inference on the true sepsis incidence is uncertain and should be interpreted with caution. It is also possible that patients’ characteristics, such as organ dysfunction and source of infection, may be different for sepsis developing in these wards. Yet, in the ICU, documentation is usually both extensive and of good quality and a similar surveillance system has performed well in this setting.49 The algorithm also showed lower sensitivity for HO sepsis compared with CO sepsis. This was primarily due to organ dysfunction only mentioned in free text, which indicates that improved recording of oxygen therapy and vital parameters such as GCS could result in better algorithm performance in surveillance of HO sepsis. Moreover, an implemented surveillance system requires continuous maintenance and validation. Although we used an exact duplicate of the EHR system, our algorithm has not yet been implemented and also needs evaluation in a real-world scenario. Finally, the study was limited to a single centre and needs confirmation within different EHR systems in different hospitals.

**Conclusion**
Based on data from EHR, it is feasible to automatically monitor sepsis incidence with good validity compared with physician medical record review in non-intensive care wards using the Sepsis-3 clinical criteria as surveillance definition. The algorithm exposed variations in hospital-onset sepsis incidence depending on ward type, which can be used to tailor infection prevention interventions and improve sepsis care.

**Author affiliations**
1Division of Infectious Diseases, Department of Medicine, Solna (MedS), Karolinska Institutet, Stockholm, Sweden
2Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden
3Treat Systems ApS, Aalborg, Denmark
4Center for Model-based Medical Decision Support, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark
5Department of Clinical microbiology and the Laboratory for Molecular Infection Medicine (MIMMS), Umeå University, Umeå, Sweden
6Department of Anesthesiology and Perioperative medicine, Mayo Clinic, Rochester, Minnesota, USA
7Department of Computer and Systems Sciences, Stockholm University, Kista, Sweden

**Contributors** Concept and design: PN, JKV, LW, VH, BP, AF, AT, AFJ, MLM, HD, AH. Acquisition, analysis, or interpretation of data: JKV, LW, PN, EA, HT, KM, MLM, HD, AH. Drafting of the manuscript: JKV, PN, LW. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: LW, JKV, PN. Obtained funding: PN, AF, AFJ, MLM, HD, JKV, BP, VH.

**Funding** The work was supported by Vinnova (grant 2016-00563). JKV was supported by Region Stockholm (combined clinical residency and PhD training program). PN was supported by Region Stockholm (clinical research
appointment). JKV, PN, VH and BP received the Mayo Clinic-Karolinska Institutet Collaborative Travel Award 2017.

**Competing interests** LW and MLM are employees of Treat Systems ApS (Aalborg, Denmark). Treat Systems produces medical decision support systems for antimicrobial and microbiological diagnostic stewardship.

**Patient consent for publication** Not required.

**Ethics approval** The study was approved by the Regional Ethical Review Board in Stockholm under permission no. 2016/2309-32 and 2012/1838-31/3

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data from deidentified electronic health records are not freely available due to protection of the personal integrity of the participants. Access to patient level data requires a Swedish ethical permit and an agreement with the research organisation, Department of Computer and Systems Sciences, Stockholm University, holder of the data.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID iDs**

John Karlssom Valik http://orcid.org/0000-0002-4521-1886

Logan Ward http://orcid.org/0000-0002-2697-3922

Anders P Johansson http://orcid.org/0000-0003-0548-594

**REFERENCES**


