

Enhanced Screening and Research Data Collection via Automated EHR Data Capture and Early Identification of Sepsis

SAGE Open Nursing
Volume 5: 1–11
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2377960819850972
journals.sagepub.com/home/son



Reba Umberger, PhD, RN, CCRN-K¹ ,
Chayawat “Yo” Indranoi, MSIE², Melanie Simpson, BSN, CCRN²,
Rose Jensen, BSN², James Shamiyeh, MD, FCCP², and
Sachin Yende, MD, MS³

Abstract

Clinical research in sepsis patients often requires gathering large amounts of longitudinal information. The electronic health record can be used to identify patients with sepsis, improve participant study recruitment, and extract data. The process of extracting data in a reliable and usable format is challenging, despite standard programming language. The aims of this project were to explore infrastructures for capturing electronic health record data and to apply criteria for identifying patients with sepsis. We conducted a prospective feasibility study to locate and capture/abstract electronic health record data for future sepsis studies. We located parameters as displayed to providers within the system and then captured data transmitted in Health Level Seven[®] interfaces between electronic health record systems into a prototype database. We evaluated our ability to successfully identify patients admitted with sepsis in the target intensive care unit (ICU) at two cross-sectional time points and then over a 2-month period. A majority of the selected parameters were accessible using an iterative process to locate and abstract them to the prototype database. We successfully identified patients admitted to a 20-bed ICU with sepsis using four data interfaces. Retrospectively applying similar criteria to data captured for 319 patients admitted to ICU over a 2-month period was less sensitive in identifying patients admitted directly to the ICU with sepsis. Classification into three admission categories (sepsis, no-sepsis, and other) was fair (Kappa .39) when compared with manual chart review. This project confirms reported barriers in data extraction. Data can be abstracted for future research, although more work is needed to refine and create customizable reports. We recommend that researchers engage their information technology department to electronically apply research criteria for improved research screening at the point of ICU admission. Using clinical electronic health records data to classify patients with sepsis over time is complex and challenging.

Keywords

severe sepsis, critical care, electronic health records (EHR), Health Level Seven (HL7)

Date received: 22 December 2018; revised: 1 March 2019; accepted: 20 April 2019

Introduction

Sepsis—“a life-threatening organ dysfunction caused by a dysregulated host response to infection”—is a significant public health concern and major contributor to morbidity and mortality (Singer et al., 2016, p. 801). Approximately 19 million people develop sepsis worldwide each year, one in three die during the year after hospitalization, and one in six have persistent impairments (e.g., physical, cognitive, psychological, and

¹Department of Acute and Tertiary Care, College of Nursing, The University of Tennessee Health Science Center, Memphis, TN, USA

²University Health System, The University of Tennessee Medical Center, Knoxville, TN, USA

³Department of Critical Care Medicine, University of Pittsburgh, PA, USA

Corresponding Author:

Reba Umberger, Department of Acute and Tertiary Care, College of Nursing, The University of Tennessee Health Science Center, 920 Madison Avenue, Suite 551, Memphis, TN 38163, USA.

Email: rumberge@uthsc.edu



immune dysfunction; Prescott & Angus, 2018). Prospective cohort studies and clinical trials in sepsis (e.g., Protocolised Management In Sepsis [Mouncey et al., 2015], Protocolized Care for Early Septic Shock [Yealy et al., 2014], Australasian Resuscitation in Sepsis Evaluation [Delaney et al., 2013], and Sepsis survivors MOnitoring and coordination in Outpatient health care [Schmidt et al., 2014; Schmidt et al., 2016]) require identifying sepsis patients and collecting large amounts of data for each participant over time. Ascertaining long-term outcomes to evaluate causal relationships in critical illness and sepsis require careful collection of exposures that occurred during the intensive care unit (ICU) stay throughout the follow-up period (Needham et al., 2012; Needham, Dowdy, Mendez-Tellez, Herridge, & Pronovost, 2005; Yende & Angus, 2007). Cohort studies have the advantage of allowing researchers to examine the temporal relationships of multiple exposures to outcomes, even those that are potentially harmful (Dowdy, Needham, Mendez-Tellez, Herridge, & Pronovost, 2005). Data collection can be costly in terms of time and personnel. Study personnel must collect each data point from a patient's medical record, transcribe data onto paper forms, enter data into a database, and perform quality assurance/data cleaning, all of which are time consuming, introduce errors, and sometimes cost prohibitive. An alternative is to directly abstract data from electronic health records (EHR), but this approach has challenges. We will first review literature related to data abstraction from the EHR.

Review of Literature

The use of EHR for research has become more frequent over the past decade. There has been a trend to move away from simply using administrative data sets (e.g., billing/claims data) to more fully engaging the capabilities of EHR for outcomes research. Dean et al. (2009) were the first to systematically review the use of EHR systems in the United States for health outcomes research. The authors defined health outcomes broadly (e.g., comorbidities, risk factors, medical care utilization, diagnostic testing, patient-reported data, adverse events, and costs) in their review of 126 studies. They discussed several issues pertaining to the quality of EHR data and suggested that researchers become familiar with the nuances of their EHR system and "the degree to which data fields accurately and comprehensively capture patient care" (p. 626). Although EHR is standard for documenting clinical care, variations across systems may limit research use. The degree of validation needed for a research study will vary based on the proposed research to be conducted. One major issue is a lack of standardization in terms used by providers. There are no standard methods for retrieving and mining EHR data,

although implementation and use of standard terminology (e.g., Systematized Nomenclature of Medicine-Clinical Terms, National Health Information Network, and others) are helping to reduce variability for data coding and capture (Dean et al., 2009; Jensen, Jensen, & Brunak, 2012). The use of EHR data is subject to selection bias and confounding. Some studies that controlled for potential selection bias and confounding primarily used multivariate regression analysis, some used matching and stratification, and only one used propensity score methods. It was observed that many studies also required collection of supplemental data (e.g., patient reports such as quality of life surveys), and it was suggested that integration of such data into future EHR systems could improve assessments by providers and researchers (Dean et al., 2009).

A more recent review by Lin, Jiao, Biskupiak, and McAdam-Marx (2013) identified 96 articles using EHR for research that includes health outcomes. They describe common research issues pertaining to data location and format. For example, free-text data may be difficult to locate and retrieve. One way to overcome this problem is with the use of Natural Language Processing (NLP) to search for key terms within text documents. Only two studies in their review used this approach. One examined recurrent depression and the other examined records for postoperative complications. NLP will be an important method for future research, but much more pioneering work from experts in the field will be needed to design and validate NLP concepts used to assess outcomes and exposures (Lin et al., 2013). The external validity of any research findings using EHR data depends on the experimental design and accuracy of the data retrieved. EHR, even different versions from the same developer, differ because they are customizable to meet the needs to each facility and specific departments within a facility.

EHR systems are designed for clinicians at the point of care and are not designed for research. Thus, there is often a redundancy built into the systems for ease of clinical use by allowing multiple locations for entering and retrieving information (Terry et al., 2010). Terry et al. (2010) categorized 285 abstracts of EHR studies in primary care and provided a "primer" to help researchers planning to engage in EHR research in avoiding pitfalls. They identified few studies focused on data quality and even fewer focused on ethics and privacy in using EHR data for research. The authors shared five helpful considerations for researchers planning to use EHR data: (a) data may be entered by providers in various locations, (b) data may be entered in various formats, (c) data may be entered by providers using inconsistent terms, (d) data may not be readily searchable, and (e) data not required for clinical care may be missing (Terry et al., 2010). They also discussed five

levels of data extraction, with each level increasing in complexity. The first three are all data queries: (a) pre-determined, (b) customizable, and (c) advanced customizable. The EHR system usually allows for users to generate these queries with more advanced queries using Boolean logic. The last two are as follows: (d) structured query language (SQL) interface and (e) data extraction and analysis with database tools. These higher level forms of data abstraction require collaboration with providers, researchers, and information technology professionals. Researchers will need the EHR's entity relationship diagram to understand the relationships among the EHR data files (Terry et al., 2010).

Many challenges have been identified in regard to using EHR data for research purposes. These challenges also include processes to assure the availability of accurate and valid data needed for the specific research project as well as the use of safeguards and privacy in data mining (Jensen et al., 2012). The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 encourages the use and development of health information technology—improving health by making the EHR accessible to care providers, researchers, and public health workers (Blumenthal, 2010). As health information technology is more uniformly adopted, terminology becomes standardized, and systems become more integrated, so that research using EHR will allow linkage of information to advance care and provide individualized patient-centered precision medicine (Collins & Varmus, 2015). Currently, research involving EHR data requires careful review of the data available while assessing the ability to capture that data electronically and validating the data.

Purpose

The authors conducted a feasibility study to prospectively abstract EHR data for sepsis patients. The goals of this study were to (a) explore infrastructure for capturing data, (b) identify medical intensive care patients admitted with sepsis, and (c) validate our ability to correctly identify patients with sepsis electronically based on specific criteria.

Methods

This prospective feasibility study examined the local infrastructure for capturing EHR data, applied coding rules and restrictions to electronically identify patients in the medical ICU (MICU) with sepsis, captured EHR data from data streams (described below), and validated automated sepsis classification in two small cross-sectional samples and one large sample gathered or “captured” over 2 months. This study was approved by the university and hospital institutional review boards.

Individual informed consent was waived by the institutional review board. All identifiable EHR data for this project remained behind the hospital firewall, accessible only to the research team. Methods to identify the study population, role of the team, data collection, and data analysis follows.

Study Population

The target population was patients admitted with sepsis to a 20-bed MICU in an academic southeast U.S. medical center. Initially, our focus was the combination of EHR data fields, variables, and infrastructures, rather than patients. Next, we identified sepsis patients using the following automated EHR programmed criteria:

- Admitted directly to the medical intensive care unit
- Elevated white blood cell count >12,000 or <4,000
- Temperature > 38.3°C or < 36°C
- Receiving antibiotics. We restricted to common antibiotics used for sepsis patients, including Piperacillin/Tazobactam, Moxifloxacin, Ceftriaxone, Clindamycin, Azithromycin, Gentamycin, and Vancomycin.
- Admission history and physical data that used the following terms: sepsis, septic shock, pneumonia, community-acquired pneumonia, health-care-associated pneumonia, bacteremia, urosepsis, and urinary tract infection.

These criteria are similar to the criteria used to develop the Sepsis-3 definitions using EHR data as well as prior definitions (Levy et al., 2003; Shankar-Hari et al., 2016; Singer et al., 2016). We sought to identify patients with suspected sepsis to facilitate additional screening by study personnel prior to seeking informed consent. Individual chart review was targeted to classify patients into three groups, namely: Group 1: patients who were not admitted directly to the ICU, who were in the ICU for less than 48 hours, or were younger than 18 years old; Group 2: directly admitted to ICU with sepsis; and Group 3: directly admitted to the ICU without sepsis. Specifically, we manually reviewed identified EHR for the following inclusion criteria:

- ICU length of stay greater or equal to 48 hours
- Age greater than 18 years
- Direct MICU admission
- Possible sepsis code (based on chief complaint and history and physical)
- A positive Systemic Inflammatory Response Score (SIRS) at admission (at least 2 of 4)
- Individual positive or negative SIRS components (white blood cell, Temperature, heart rate, respiratory rate)

- The presence of antibiotics at admission
- Positive cultures at admission

Exclusion criteria included:

- Patients who were not admitted directly to the ICU
- Patients who were in the ICU for less than 48 hours
- Patients younger than 18 years old

When exclusion criteria were identified, full screening was not collected. These excluded patients were classified as Group 1, the excluded group.

Study Team

We used a multidisciplinary team. The principal investigator, R.U., has a background in critical care and sepsis research, C.C. had worked as a nurse in MICU and routinely used the Cerner Millennium® EHR (Cerner©2016, Cerner Corporation, North Kansas City, MO), and C.I. is a data systems analyst with expertise in programming and working with the EHR systems. R.U. developed the list of variables to assess (Table 1) and worked with C.C. to locate each variable within the EHR system. They communicated these details to C.I., who used data mining techniques to assess our ability to capture the variables in Table 1 and to electronically identify patients admitted to the MICU with sepsis using the criteria identified earlier. M.S. assisted with initial cross-check validation. M.S. was an MICU Team Leader who assisted with chart reviews. R.U. and M.S. reviewed automatic classification from the prototype against clinical records in Cerner as noted above to determine group classification. R.J. was MICU Nurse Manager and assisted with early conceptualization and understanding the unit's existing work in sepsis surveillance. J.S. was the MICU Comedical Director and provided consultation for later phases of the project.

Data Collection

The goals of this study were to (a) explore infrastructure for capturing data, (b) identify MICU admitted with sepsis, and (c) validate our ability to correctly identify patients with sepsis based on specific criteria. First, we developed a spreadsheet with a list of parameters (Table 1) to extract from the Cerner database. To explore the infrastructure for capturing data, we reviewed the front-end user display and provided the analyst with our observed location of each parameter. We tested accessibility of the data and created an infrastructure to capture the data for reports while not interfering with EHR performance for front-end users. We created a multidimensional data set and flat tables for data visualization and aggregation, a prototype database

(PDB) as shown in Figure 1. All of these data remained protected behind the hospital firewall. We manually compared accessibility of each parameter—as displayed within the PDB and as displayed to front-end users in Cerner—on the spreadsheet until we observed nothing new (had reached saturation). We provided feedback to the analyst for adjustments to the PDB as needed.

Next, to identify patients in the MICU with sepsis at two time points, the analyst applied sepsis criteria (described under study population) to the list of patients on the MICU census while two nurses reviewed the census and made visual comparisons. To validate our ability to correctly identify patients with sepsis, we conducted individual chart review and compared manual review with computer classifications applied within the PDB. We used approaches used previously (Figure 1) to capture data using Health Level Seven (HL7), a standard language for integrating and exchanging electronic health data to mediate the flow of information from various hospital systems (Health Level Seven® International). For example, we used the following HL7 message types between interfaces to capture health system data streams:

- ADT (Admission, Discharge, Transfer) for demographic and coding information;
- ORU (Observation Result) for laboratory results, vital signs;
- RDE (Pharmacy/Treatment Encoded Order) for medication/antibiotics information; and
- MDM (Medical Document Management) for transcription documents like H&P.

SQL was used to extract data from Cerner Millennium® (health information system in Figure 1). We pulled parameters (target variables described in Table 1) and used filters and rules (as detailed in the inclusion criteria above) to identify patients with sepsis. We also tested methods to de-identify the data by assigning a unique identification number which is meaningless to the Cerner Millennium® database and cannot be used to search for patient information outside of the PDB.

Data Analysis

We made visual comparisons to examine data presented in the PDB with the data displayed within the Cerner EHR system. We evaluated individual records and made comparisons until we reached saturation of the findings for variables in Table 1 ($n=5$ EHR records) when examining the ability to capture a selected variable. Cross-sectional verification also did not require data analysis. The analyst applied filters and codes to identify patients with sepsis to generate a list of patients with sepsis at two time points. Two nurses compared that

Table 1. Phase I: Variables Assessed for Ability to Capture Electronically Behind the Hospital Firewall.

Type of measure	Variables	Frequency	Data system
Demographics	Age, sex, race, height, and admission weight	Single measure	HIS
Baseline variables	Primary reason for hospital admission, primary reason for ICU admission, past medical history, ICU admission, and Admission APACHE II Score.	Single measure	HIS, MRS
Outcome variables	Hospital admission date, hospital discharge date, calculate hospital length of stay, discharge destination, ICU admission date, ICU hospital discharge date, ICU discharge date, calculate ICU length of stay, ICU destination, primary discharge diagnosis, any return to the ICU within the same hospital admission, list of all discharge diagnoses (ICD-9 or 10), and procedure codes.	Single measure and multiple measures	HIS, MRS
Laboratories	All cultures results and select <i>daily</i> laboratories (Total Bilirubin, Creatinine, Platelet, Hematocrit, White Blood Cell Count, % Bands [if present], Glucose, Amylase, Lipase, Albumin, C-reactive protein, Lactate, and Arterial Blood Gas Results [with O ₂ setting associated with the results])	Multiple measures	HIS, LIS, POCT
Vital signs	All recorded vital signs (Systolic Blood pressure, Diastolic blood pressure, mean arterial pressure, heart rate, respiratory rate, temperature (w/route), and Oxygen saturation with specific output times).	Multiple measures, daily (0800, minimum, and maximum values)	HIS
Ventilator/oxygen	Oxygen delivery, Ventilator settings (if present; mode, oxygen concentration, tidal volume, PEEP, pressure support, minute volume, and static compliance)	Multiple measures (0800 and 2000 settings)	HIS, POCT
Invasive devices	The presence of peripheral IVs, foley catheters, central lines, arterial lines, endotracheal tubes, tracheostomy tubes, drains/tubes, and so on.	Multiple measures	HIS
Pharmacy records	All antibiotics (date and time of first and last dose received, ordered dosage, and frequency), steroids, and vasopressor use.	Multiple measures	HIS, PIS
SIRS score	SIRS calculated with Sepsis Alert notifications.	Multiple measures	HIS
Nutrition	Albumen level (if present) and type of diet	Multiple measures	HIS
Other	Date and type of any blood products received and the time and date of any blue alert (code).	Multiple measures	HIS

Note. APACHE II is composite measure of the highest severity of illness, composed of scores representing minimum or maximum physiologic variables within a 24-hour period, as well as scores for age and specific chronic illnesses. ICD = international classification of disease; PEEP = positive end expiratory pressure; HIS = health information system; MRS = medical record system; POCT = point-of-care testing system; PIS = pharmacy information system; LIS = laboratory information system; APACHE II = acute physiologic and chronic health evaluation score, version 2; SIRS = Systemic Inflammatory Response Score.

list with the list of patients with sepsis on the unit census. Next, longitudinal data captured in the PDB over 2 months were classified into the three groups as described earlier. Kappa statistics were calculated comparing automatic computer classifications with chart review classifications.

Results

Ability to Abstract EHR Data

A PDB was successfully created behind the hospital firewall. The data were displayed in a series of tables through a password-protected portal. The team was provided with a link to the portal for ease of access within

the hospital's intranet. One table included identifiable information (i.e., each patient's medical record number, a linking identification number, hospital and ICU admission and discharge dates, calculated hospital and ICU length of stay, chief complaint, name, gender, age, and DOB) to enable comparisons of data within the hospital EHR against the data that were captured and displayed in the PDB tables. A majority of the variables in Table 1 were accessible using an iterative process to locate them. There were three versions of the PDB over the course of the entire project. One important finding was that corrected laboratory values (incorrect values that were corrected by the laboratory) were not identified using this approach. We only identified them by examining data for inconsistencies in values that were beyond normal range.

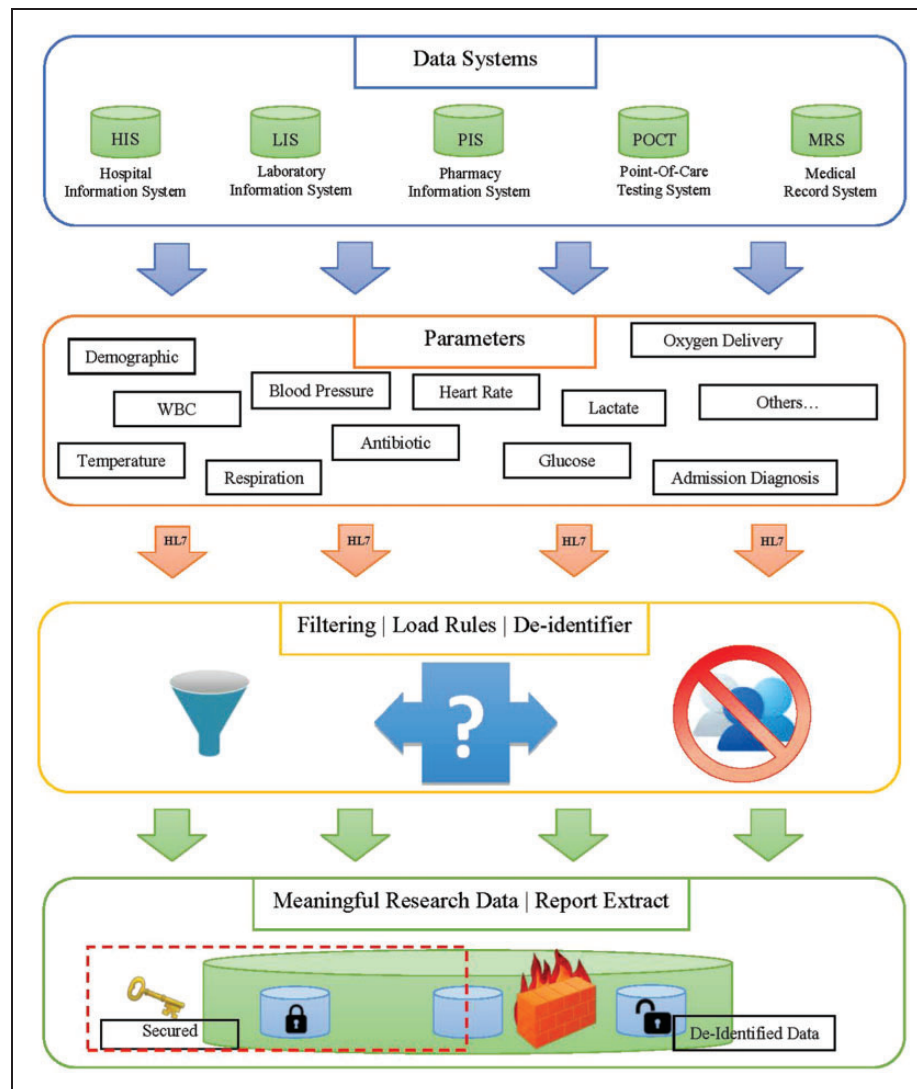


Figure 1. Conceptual data flow design. The methods used to abstract data from health systems to a prototype database (PDB) behind the hospital firewall are depicted. HL7 denotes Health Level Seven[®] International interface language. A password-protected portal to the PDB allowed team access to review and compare captured data to health system data. Later phases of this project explore secure methods of extracting de-identified data from the PDB for statistical analysis. HIS = health information system; LIS = laboratory information system; PIS = pharmacy information system; POCT = point-of-care testing system; MRS = medical record system; WBC = white blood cell.

Many of the data points were not easily accessible from either front-end applications (such as the nursing Kardex, which is a custom hospital program) or back-end queries (from the Cerner Database and requiring Oracle Procedural Language/SQL, Cerner Command Language, and Cerner's Visual Developer Tool). For example, data points that were difficult to access that appeared on the nursing Kardex but were accessible with additional coding were ICU admission and discharge dates and ICU discharge destination. Data points that were challenging for a variety of reasons included laboratory values, mechanical ventilation settings, varying location of vital sign parameters, and medications. Data points that were not accessible were

invasive devices, blood products, code blue, and diet. Free-text documents such as the H&P were accessible as a complete document for review, but we did not use NLP to pull or locate details from this document and reviewed them in context. Culture results were also formatted as Cerner encrypted text but were not accessible within the PDB.

Testing Computer Classification of Patients With Sepsis in Cross-Sectional Samples

We performed two cross-sectional evaluations of our ability to identify patients with sepsis who were in the ICU at that time. During the first assessment, we

successfully electronically identified 5 of 6 patients with sepsis (among 19 in the target unit). One patient with COPD was mistakenly identified as having sepsis by computer programming possibly related to antibiotic use and respiratory rate. One patient with septic endocarditis was identified by computer programming and initially missed by chart review until further electronic documents were manually reviewed. Programming changes were made prior to the next assessment. During the second assessment, all patients who had sepsis (among the 17 patients in the target unit) were correctly identified as having sepsis and two patients were incorrectly identified as having sepsis (false positive). One of those patients had an infection, but no SIRS and the other patient had an altered mental status for reasons other than severe sepsis. Altered mental status is a frequent occurrence in sepsis, so these terms were included in the chief complaint evaluation by the computer program. Patients with sepsis were successfully identified using four data interfaces (health information system, laboratory information system, pharmacy information system, and medical record system).

Testing Computer Classification of Patients With Sepsis in Longitudinal Data

As described earlier, a relational PDB was developed behind the hospital firewall to collect data streams from hospital system interfaces over a 2-month period. Table 2 shows computer classification and chart review classification among 319 patients admitted during this period. Percentage agreement was 87.5%, 16.7%, and 59.0% between computer and chart classification among chart confirmed cases with 208, 72, and 39 cases classified in Group 1 (patients who were not admitted directly to the ICU, who were in the ICU for

less than 48 hours, or were younger than 18 years old), Group 2 (directly admitted to ICU with sepsis), and Group 3 (directly admitted to the ICU without sepsis), respectively. Although the overall table agreement was fair (Kappa .39), the groups of most interest (sepsis/suspected sepsis at ICU admission) had poor agreement. Feedback was given to the programmer to improve classification/coding to identify sepsis patients. Classification using a single time point (upon admission to the ICU) was more accurate than classifying patients based on data over time due to the changing nature of many classification variables over time (e.g., antibiotics).

Discussion

This feasibility project was an iterative process requiring identification of data streams to capture specific data points (parameters), verify our ability to correctly and electronically identify patients with sepsis by manual EHR review, and refine parameter search terms until parameters and patients with sepsis were accurately identified. Although the HL7 standard is widely used in health care, existing barriers remain in seamlessly extracting data from the EHR in an analyzable format.

Abstraction Data From the EHR

Clinical EHR data, like administrative data sets, are not created for research purposes and present challenges (Hersh et al., 2013). This project was designed to test the ability to pull data from the EHR specifically for the purposes of future research and to supplement directly collected bedside information in experimental and nonexperimental research (from consenting patients). The majority of data abstraction in this feasibility study required higher level abstraction techniques (SQL and above) as described by Terry et al. (2010). The

Table 2. Classification of EHR Data for 319 MICU Patients Admitted Over a 2-Month Period.

Classification	Defined	Automatically classified	Chart review classification	Agreement
Group 1	Not directly admitted to MICU from the ER or admitted for less than 48 hours	209 (65.5%) ^a	208 (65.2%)	182 (87.5%)
Group 2	Directly admitted to MICU with sepsis or suspected sepsis	16 (5%)	72 (22.6%)	12 (16.7%)
Group 3	Directly admitted to MICU without sepsis at admission	94 (29.5%)	39 (12.2%)	23 (59.0%)

Note. The Kappa statistic of .39 (.32–.46) comparing automated versus manual classifications. This statistic indicates fair agreement, but these results are driven by the large number in Group 1. Considering chart review as the gold standard, percentage agreement is shown between computer classification and chart review in the third column. The agreement is poor when comparing the ability to distinguish between those with and without sepsis at admission. This may be explained by chart review allowing for a detailed review of the H&P, Systemic Inflammatory Response Score, antibiotics, and cultures at the time of admission. MICU: medical ICU; ER: emergency room; EHR: electronic health records.

^aTen patients had missing records and could not be automatically classified. They were excluded (Group 1).

process of creating the PDB to access selected parameters required a team and several iterations, which is consistent with other projects evaluating the use of EHR for research (Apte et al., 2011). The team members spent extensive time validating the data, and a rapid PDB visualization tool was useful during the design sessions/team meetings. Most of the challenges we detected have been experienced by others.

Parameters location. There are multiple ways that nurses can document vital signs as well as other parameters. Data points for the same laboratory test may originate from a point-of-care system or the laboratory information system. Further, heart rate can be abstracted from an EKG monitor, pulse oximeter, or manually entered. In addition, vital signs may reside in fields specific to ICU vital signs rather than routine vital signs taken in a ward. It is challenging when all components of calculated values are not present at the same time. For example, SIRS scores include four components, but each component is not collected with the same frequency or timing. Capturing the vital signs data was one of the most challenging parts of the project. It is important for clinical researchers and IT to work together to specify which variables are to be used for research, and protocol congruent rules should be put in place for handling missing data to prevent statistical imputations (that may regress to the mean) when real data (possibly within even a few minutes) are available.

Multiple parameter measures. The problem of which values to select when there are multiple measures taken on the same day is not unique to EHR studies. Operational designations need to be made in advance and clearly specified in clinical trial protocols. For example, studies that examine longitudinal data may collect data at a single time point each day (e.g., 8:00 a.m.), and in so doing miss events that occur at other time points. One way we accounted for variability is to capture it by recording daily minimum and maximum values for parameters, in addition to a standard time. We were able to generate this information in PDB reports. Studies that examine only one time point may miss important variability.

Parameter format. As mentioned in the “Introduction” section, clinicians often document in more than one location, and free-text notes allow them flexibility to express case details. Free text is commonly used, is heterogeneous, and can be challenging to analyze. NLP and machine learning techniques are used to extract text data (Jensen et al., 2012). For the purpose of our project, we pulled full reports which could be viewed in context by researchers in prospective experimental or nonexperimental bedside research.

EHR system changes. EHR updates can be required by institutions (programmed internally) to allow for capture of Core Measures that institutions must report or based on unit specific needs. This may lead to differences in the way that providers document or differing internal methods of capturing the data. Change can also arise during system upgrades and updates from the makers of EHR systems or due to institutional decisions to add or remove packages (e.g., during this project an add-on package that enabled daily APACHE II calculation was not repurchased). It is unknown how often EHR changes occur—internally or externally—but technology is rarely static. Investigators designing projects designed to collect data from the EHR will be impacted by changes and should work closely with IT to review changes frequently.

Electronic Detection of Patients With Sepsis

Our second cross-sectional review was more sensitive, as it identified more patients as having sepsis than actually had sepsis, an increase in false positives. We could have further adapted the rules, but we preferred to have more false positives in order to prevent missing any potential patients who might have sepsis. Although we only evaluated our ability to identify patients with sepsis for screening, the next step would be to establish an automatic notification system for research. Others have electronically identified patients using investigator-designed algorithms to screen for recruitment and have thereby increased their recruitment efficiency and the sensitivity of identifying patients (Cardozo, Meurer, Smith, & Holschen, 2010; Herasevich, Pieper, Pulido, & Gajic, 2011). Cordozo et al. (2010) increased the sensitivity of screening from 5.9% to 100% after implementing automated electronic record screening versus the prior method of patients screening by paging physicians. Herasevich et al. (2011) doubled research participant recruitment from four patients per month to eight patients per month after implementation of the automated “septic shock sniffer” to identify septic shock patients using the EHR. Although their study coordinators had access to all aspects of the EHR and had used it for screening, they often used physician notes for their preliminary evaluation and their review of physiologic variables may have been limited (Herasevich et al., 2011). Time and motion studies have shown reduced screening time per patient by using an electronic screening tool—reduced from 18 minutes to less than 3 minutes (Thompson, Oberteuffer, & Dorman, 2003). Developing and using automated systems within the EHR to identify potential study participants should become a more commonplace in sepsis studies. Increased recruitment efficiency will shorten the time needed to complete clinical research (experimental

and nonexperimental studies) and will allow for more timely dissemination of findings.

The international classification of diseases (ICD) is commonly used to identify research records. These codes are typically not available at the point of care for most institutions, as they are assigned upon discharge by medical record coders based on provider assessments and specific disease criteria. These codes were designed for administrative purposes rather than research, but validated methods are available for retrospective research to identify patients with sepsis (Fleischmann-Struzek et al., 2018; Iwashyna et al., 2014; Jolley et al., 2015). Some institutions assign ICD codes upon admission and discharge, so using ICD coding may also be beneficial in prospective research in some settings.

Hospital-Wide Sepsis Alert Implementation

Sepsis is an important problem and forward-thinking institutions put systems in place to help them improve prevention, early detection, and treatment of sepsis. During this project, the hospital implemented a sepsis detection system called the St. John Sepsis Agent[®], which “crawls/iteratively searches” through the EHR and sends alerts to nurses when SIRS and organ failure criteria are met (Amland & Hahn-Cover, 2016; Am-land, Haley, & Lyons, 2016). We reported details of the early alert system’s implementation with a focus on the SIRS and MODS details among the patients in Table 2 who were admitted directly to MICU with and without sepsis (Umberger, Indrani, Simpson, Jensen, & Shamiyeh, 2013).

As the alert system was implemented, trigger parameters were modified to reduce the number of false-positive alerts. Each alert required contacting providers. Issues related to work flow following hospital-wide implementation of similar systems have been reported (Guidi et al., 2015), and the hospital did experience some of these workflow issues. J.S. spent a year facilitating meetings throughout the hospital to improve local implementation of the system (e.g., addressing false positive alerts, improving work flow-related issues, and preventing desensitization with alerts), as well as developing a sepsis bundle process. A sepsis coordinator was hired to oversee the sepsis program with the medical director. Among other duties, this coordinator is responsible for monitoring and improving the SEP-1 Core Measure performance (Drake, 2015). The development of a specific sepsis initiative led to the creation of a performance improvement data infrastructure focused on sepsis metrics. The sepsis team encountered similar challenges as those outlined in this article in terms of extracting correct and relevant information from the EHR. This suggests that any further work on processes to reliably extract sepsis data from EHRs for research purposes may

simultaneously have a positive impact on hospital performance improvement activities.

Limitations

In our project, we used data streams and evaluated the ability to abstract a broad range of data types over time. We also examined our ability to capture variability (minimum and maximum range) for several important variables that are used to evaluate severity of illness via tools like APACHE. We did not apply our review in a real-time fashion, nor did we define the sensitivity and specificity of our approach during this pilot feasibility study.

We used the SIRS criteria for this project because the SIRS criteria were part of the most recent consensus sepsis criteria at the time of implementation of this project (Levy et al., 2003). This study began before Sepsis-3 and qSOFA criteria were released (Singer et al., 2016).

Implications for Practice

Electronic methods may help facilitate earlier sepsis identification by using a combination of NLP and risk stratification tools to refine sepsis alerts using smart algorithms that consider the heterogeneity and diversity of this population. Sepsis alert systems can have high sensitivity (93%) and specificity (98%), yet still have a low positive predictive value (21%; Alsolamy et al., 2014). Implementation of automated alert systems must be designed with careful attention to work flow and prevention of potential alarm fatigue. The development of artificial intelligence and deep machine learning will allow for real-time data stream monitoring to better predict sepsis in the future (Kamaleswaran et al., 2018; Nemati et al., 2018). Real-time data gathering with precise data displayed for critical decision support will be needed for clinicians. Researchers working with IT professionals can readily identify patients with sepsis with minimal programming; however, data abstraction for research is more time consuming and programmer intensive. As hospitals and outpatient clinics become more connected, these methods may facilitate more complete long-term data collection, thus helping to reduce fragmentation of care after sepsis.

Conclusions

Although providers are able to easily view EHR data for clinical care, the abstraction of data directly from EHR systems for research purposes remains a challenge. Despite limitations, improving direct data capture methods can assist in targeting potential clinical trial participants and reduce the burden of data collection. More research is needed to determine the best methods for automatically identifying patients with sepsis.

Authors' Note

The aspects of this project have been presented at (a) Sigma Theta Tau Gamma Chi Chapter's Research Day in Knoxville, TN, (b) the University of Tennessee Medical Center in Knoxville, TN, and (c) the Society of Critical Care Medicine in San Francisco, CA.

Acknowledgments

The authors would like to acknowledge the support of Ms. Camille Coleman, MSN, RN, CCRN, for her assistance with the project as a part of a research immersion course. She assisted the R.U. with locating data within Cerner and the cross-sectional assessments. J.S. was the Medical Critical Care (MCC) Medical Director, R.J. was the MCC Nurse Manager, and M.S. was the MCC Team Leader at the time this project was designed and conducted. The authors thank Dr. Richard Redfeard (Director of Scientific Writing, Office of Research, University of Tennessee Health Science Center) for editing the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Health Information Technology and Simulation Research Unit at the University of Tennessee, Knoxville, TN.

ORCID iD

Reba Umberger  <https://orcid.org/0000-0002-4722-4031>

References

- Alsolamy, S., Al Salamah, M., Al Thagafi, M., Al-Dorzi, H. M., Marini, A. M., Aljerian, N.,... Arabi, Y. M. (2014). Diagnostic accuracy of a screening electronic alert tool for severe sepsis and septic shock in the emergency department. *BMC Medical Informatics and Decision Making, 14*, 105. doi:10.1186/s12911-014-0105-7
- Amland, R. C., & Hahn-Cover, K. E. (2016). Clinical decision support for early recognition of sepsis. *American Journal of Medical Quality, 31*(2), 103–110. doi:10.1177/1062860614557636
- Amland, R. C., Haley, J. M., & Lyons, J. J. (2016). A multi-disciplinary sepsis program enabled by a two-stage clinical decision support system: Factors that influence patient outcomes. *American Journal of Medical Quality, 31*(6), 501–508. doi:10.1177/1062860615606801
- Apte, M., Neidell, M., Furuya, E. Y., Caplan, D., Glied, S., & Larson, E. (2011). Using electronically available inpatient hospital data for research. *Clinical and Translational Science, 4*(5), 338–345. doi:10.1111/j.1752-8062.2011.00353.x
- Blumenthal, D. (2010). Launching HITECH. *New England Journal of Medicine, 362*(5), 382–385. doi:10.1056/NEJMp0912825
- Cardozo, E., Meurer, W. J., Smith, B. L., & Holschen, J. C. (2010). Utility of an automated notification system for recruitment of research subjects. *Emergency Medicine Journal, 27*(10), 786–787. doi:10.1136/emj.2009.081299
- Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795. doi:10.1056/NEJMp1500523
- Dean, B. B., Lam, J., Natoli, J. L., Butler, Q., Aguilar, D., & Nordyke, R. J. (2009). Review: Use of electronic medical records for health outcomes research: A literature review. *Medical Care Research and Review, 66*(6), 611–638. doi:10.1177/1077558709332440
- Delaney, A. P., Peake, S. L., Bellomo, R., Cameron, P., Holdgate, A., Howe, B.,... Webb, S. (2013). The Australasian Resuscitation in Sepsis Evaluation (ARISE) trial statistical analysis plan. *Critical Care Resuscitation, 15*(3), 162–171.
- Dowdy, D. W., Needham, D. M., Mendez-Tellez, P. A., Herridge, M. S., & Pronovost, P. J. (2005). Studying outcomes of intensive care unit survivors: The role of the cohort study. *Intensive Care Medicine, 31*(7), 914–921. doi:10.1007/s00134-005-2657-6
- Drake, K. (2015). The sepsis core measures initiative. *Nursing Management, 46*(12), 33–36. doi:10.1097/01.NUMA.0000473506.31055.b6
- Fleischmann-Struzek, C., Thomas-Ruddel, D. O., Schettler, A., Schwarzkopf, D., Stacke, A., Seymour, C. W.,... Reinhart, K. (2018). Comparing the validity of different ICD coding abstraction strategies for sepsis case identification in German claims data. *PLoS One, 13*(7), e0198847. doi:10.1371/journal.pone.0198847
- Guidi, J. L., Clark, K., Upton, M. T., Faust, H., Umscheid, C. A., Lane-Fall, M. B.,... Fuchs, B. D. (2015). Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. *Annals of the American Thoracic Society, 12*(10), 1514–1519. doi:10.1513/AnnalsATS.2015.03-1290C
- Health Level Seven® International. Retrieved from <http://www.hl7.org/implement/standards/>
- Herasevich, V., Pieper, M. S., Pulido, J., & Gajic, O. (2011). Enrollment into a time sensitive clinical study in the critical care setting: Results from computerized septic shock sniffer implementation. *Journal of the American Medical Informatics Association, 18*(5), 639–644. doi:10.1136/amiajnl-2011-000228
- Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V.,... Saltz, J. H. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care, 51*(8 Suppl 3), S30–S37. doi:10.1097/MLR.0b013e31829b1dbd
- Iwashyna, T. J., Odden, A., Rohde, J., Bonham, C., Kuhn, L., Malani, P.,... Flanders, S. (2014). Identifying patients with severe sepsis using administrative claims: Patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Medical Care, 52*(6), e39–e43. doi:10.1097/MLR.0b013e318268ac86

- Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. *Nature Reviews Genetics*, *13*(6), 395–405. doi:10.1038/nrg3208
- Jolley, R. J., Quan, H., Jette, N., Sawka, K. J., Diep, L., Goliath, J., ... Doig, C. J. (2015). Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open*, *5*(12), e009487. doi:10.1136/bmjopen-2015-009487
- Kamaleswaran, R., Akbilgic, O., Hallman, M. A., West, A. N., Davis, R. L., & Shah, S. H. (2018). Applying artificial intelligence to identify physiometers predicting severe sepsis in the PICU. *Pediatric Critical Care Medicine*, *19*(10), e495–e503. doi:10.1097/PCC.0000000000001666
- Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., ... Ramsay, G. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Critical Care Medicine*, *31*(4), 1250–1256. doi:10.1097/01.CCM.0000050454.01978.3B
- Lin, J., Jiao, T., Biskupiak, J. E., & McAdam-Marx, C. (2013). Application of electronic medical record data for health outcomes research: A review of recent literature. *Expert Review of Pharmacoeconomics and Outcomes Research*, *13*(2), 191–200. doi:10.1586/erp.13.7
- Mouncey, P. R., Osborn, T. M., Power, G. S., Harrison, D. A., Sadique, M. Z., Grieve, R. D., ... Rowan, K. M. (2015). Trial of early, goal-directed resuscitation for septic shock. *New England Journal of Medicine*, *372*(14), 1301–1311. doi:10.1056/NEJMoa1500896
- Needham, D. M., Davidson, J., Cohen, H., Hopkins, R. O., Weinert, C., Wunsch, H., ... Harvey, M. A. (2012). Improving long-term outcomes after discharge from intensive care unit: Report from a stakeholders' conference. *Critical Care Medicine*, *40*(2), 502–509. doi:10.1097/CCM.0b013e318232da75
- Needham, D. M., Dowdy, D. W., Mendez-Tellez, P. A., Herridge, M. S., & Pronovost, P. J. (2005). Studying outcomes of intensive care unit survivors: Measuring exposures and outcomes. *Intensive Care Medicine*, *31*(9), 1153–1160. doi:10.1007/s00134-005-2656-7
- Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, *46*(4), 547–553. doi:10.1097/CCM.0000000000002936
- Prescott, H. C., & Angus, D. C. (2018). Enhancing recovery from sepsis: A review. *Journal of the American Medical Association*, *319*(1), 62–75. doi:10.1001/jama.2017.17687
- Schmidt, K., Thiel, P., Mueller, F., Schmuecker, K., Worrack, S., Mehlhorn, J., ... Gensichen, J. (2014). Sepsis survivors monitoring and coordination in outpatient health care (SMOOTH): Study protocol for a randomized controlled trial. *Trials*, *15*, 283. doi:10.1186/1745-6215-15-283
- Schmidt, K., Worrack, S., Von Korff, M., Davydow, D., Brunkhorst, F., Ehlert, U., ... SMOOTH Study Group (2016). Effect of a primary care management intervention on mental health-related quality of life among survivors of sepsis: A randomized clinical trial. *Journal of the American Medical Informatics Association*, *315*(24), 2703–2711. doi:10.1001/jama.2016.7207
- Shankar-Hari, M., Phillips, G. S., Levy, M. L., Seymour, C. W., Liu, V. X., Deutschman, C. S., ... Singer, M. (2016). Developing a new definition and assessing new clinical criteria for septic shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Journal of the American Medical Association*, *315*(8), 775–787. doi:10.1001/jama.2016.0289
- Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., ... Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *Journal of the American Medical Association*, *315*(8), 801–810. doi:10.1001/jama.2016.0287
- Terry, A. L., Chevendra, V., Thind, A., Stewart, M., Marshall, J. N., & Cejic, S. (2010). Using your electronic medical record for research: A primer for avoiding pitfalls. *Family Practice*, *27*(1), 121–126. doi:10.1093/fampra/cmp068
- Thompson, D. S., Oberteuffer, R., & Dorman, T. (2003). Sepsis alert and diagnostic system: Integrating clinical systems to enhance study coordinator efficiency. *CIN: Computers Informatics Nursing*, *21*(1), 22–26. quiz 27-28.
- Umberger, R., Indrani, C., Simpson, M., Jensen, R., & Shamiyeh, J. (2013). Implementation of an automated electronic medical record tool to identify patients with sepsis. *Critical Care Medicine*, *41*(12), A269.
- Yealy, D. M., Kellum, J. A., Huang, D. T., Barnato, A. E., Weissfeld, L. A., Pike, F., ... Angus, D. C. (2014). A randomized trial of protocol-based care for early septic shock. *New England Journal of Medicine*, *370*(18), 1683–1693. doi:10.1056/NEJMoa1401602
- Yende, S., & Angus, D. C. (2007). Long-term outcomes from sepsis. *Current Infectious Disease Reports*, *9*(5), 382–386. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17880848>.