Clinical Decision-Support Systems for Detection of Systemic Inflammatory Response Syndrome, Sepsis, and Septic Shock in Critically Ill Patients: A Systematic Review

Antje Wulff[1],[*]   Sara Montag[2],[*]   Michael Marschollek[1]   Thomas Jack[2]

[1] Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany
[2] Department of Pediatric Cardiology and Intensive Care Medicine, Hannover Medical School, Hannover, Germany

Address for correspondence Antje Wulff, MSc, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany (e-mail: antje.wulff@plri.de).

Abstract

Background   The design of computerized systems able to support automated detection of threatening conditions in critically ill patients such as systemic inflammatory response syndrome (SIRS) and sepsis has been fostered recently. The increase of research work in this area is due to both the growing digitalization in health care and the increased appreciation of the importance of early sepsis detection and intervention. To be able to understand the variety of systems and their characteristics as well as performances, a systematic literature review is required. Existing reviews on this topic follow a rather restrictive searching methodology or they are outdated. As much progress has been made during the last 5 years, an updated review is needed to be able to keep track of current developments in this area of research.

Objectives   To provide an overview about current approaches for the design of clinical decision-support systems (CDSS) in the context of SIRS, sepsis, and septic shock, and to categorize and compare existing approaches.

Methods   A systematic literature review was performed in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement. Searches for eligible articles were conducted on five electronic bibliographic databases, including PubMed/MEDLINE, IEEE Xplore, Embase, Scopus, and ScienceDirect. Initial results were screened independently by two reviewers based on clearly defined eligibility criteria. A backward as well as an updated search enriched the initial results. Data were extracted from included articles and presented in a standardized way. Articles were classified into predefined categories according to characteristics extracted previously. The classification was performed according to the following categories: clinical setting including patient population and mono- or multicentric study, support type of the system such as prediction or detection, systems characteristics such as knowledge- or data-driven algorithms used, evaluation of methodology,

Keywords
► clinical decision-support systems
► sepsis
► SIRS
► critical care
► systematic review

---

[*] *These authors contributed equally to the work.*

and results including ground truth definition, sensitivity, and specificity. All results were assessed qualitatively by two reviewers.

Results   The search resulted in 2,373 articles out of which 55 results were identified as eligible. Over 80% of the articles describe monocentric studies. More than 50% include adult patients, and only four articles explicitly report the inclusion of pediatric patients. Patient recruitment often is very selective, which can be observed from highly varying inclusion and exclusion criteria. The task of disease detection is covered in 62% of the articles; prediction of upcoming conditions in 33%. Sepsis is covered in 67% of the articles, SIRS as sole entity in only 4%, whereas 27% focus on severe sepsis and/or septic shock. The most common combinations of categories "algorithm used" and "support type" are knowledge-based detection of sepsis and data-driven prediction of sepsis. In evaluations, manual chart review (38%) and diagnosis coding (29%) represent the most frequently used ground truth definitions; most studies present a sample size between 10,001 and 100,000 cases (31%) and performances highly differ with only five articles presenting sensitivities and specificities above 90%; four of them using knowledge-based rather than machine learning algorithms. The presentations of holistic CDSS approaches, including technical implementation details, system interfaces, and data and interoperability aspects enabling the use of CDSS in routine settings are missing in nearly all articles.

Conclusions   The review demonstrated the high variety of research in this context successfully. A clear trend is observable toward the use of data-driven algorithms, and a lack of research could be identified in covering the pediatric population as well as acknowledging SIRS as an independent and threatening condition. The quality as well as the significance of the presented evaluations for assessing the performances of the algorithms in clinical routine settings are often not meeting the current standard of scientific work. Our future interest will be concentrated on these realistic settings by implementing and evaluating SIRS detection approaches as well as considering factors to make the CDSS useable in clinical routine from both technical and medical perspectives.

Introduction

Rationale and Background

Although the overall burden of bloodstream infections is still rather underestimated,[1] sepsis is definitely well known to be a leading cause of death all over the world.[2] Especially critically ill patients are at higher risk for the development of sepsis and thereby increased mortality.[3] For a long time, sepsis detection was intimately connected with the systemic inflammatory response syndrome (SIRS), which is defined as a nonspecific inflammatory process in the absence of infection.[4] Although this definition of SIRS in adult patients is no longer officially used according to the new Sepsis-3 definitions,[5] the occurrence of SIRS in clinical reality still predisposes patients to organ dysfunction and organ failure, and frequently determines the clinical outcomes.[6] Furthermore, recognition of sepsis and even septic shock is often very difficult, especially in newborns and infants, so that many cases of critically ill patients could be missed when suppressing SIRS.[7–9] Patients suffering from SIRS have a high risk for a longer time of stay in intensive care units (ICUs)[10] and increased morbidity and mortality, especially in the pediatric cohort.[11] SIRS carries

the risk to progress to sepsis, severe sepsis, or septic shock, which even more aggravates the hazard of mortality. An early recognition and therapy should have a high priority to prevent such outcomes,[12] especially when considering that the risk of mortality in untreated pediatric septic shock patients was reported as increasing by 40% every hour.[13] Increased mortality in septic children was reported if fluid resuscitation was under 40 mL/kg in the first hour and initiated more than 30 minutes after diagnosing septic shock.[14]

To facilitate the identification of patients at risk for SIRS, sepsis, and septic shock, several criteria and manual scoring systems exist. For adult patients, there are the aforementioned SIRS criteria,[15] Sequential Organ Failure Assessment (SOFA) score,[16] quick SOFA score,[17] APACHE IV, and the Surviving Sepsis Campaign guidelines 2016.[5] Familiar scores used in pediatric patients are SIRS criteria and sepsis definition,[4] pSOFA,[18] Pediatric Early Warning Score (PEWS),[19] and Pediatric Logistic Organ Dysfunction Score (PELODS).[20] For neonates, a consensus score is not available. However, there seems to be a de facto agreement that treatment is justified when clinically neonatal sepsis is suspected.[21]

Against the background of digitalization in health care, coming along with an increased use of electronic health records (EHRs), patient data management systems (PDMS) in critical care settings and others, the implementation of computerized systems able to support automated SIRS and sepsis detection has been fostered. The use of such clinical decision-support systems (CDSS) even might become indispensable in near future, especially in medical specialties that require a wide variety of knowledge, which is needed to grade a huge amount of unfiltered information and to make time-sensitive and accurate decisions. It can be hypothesized that being supported by a CDSS might make the difference between acting in a timely manner and preventing clinical deterioration, or being too late. Another important aspect in favor of using CDSS is a reduction of workload for medical staff.[22] Especially within intensive care settings, clinicians reach decisions under challenging conditions, characterized by time pressure, work interruptions, and life-threatening situations and multiple subspecialties are involved in the decision-making process. A very special department featured by such characteristics is pediatric critical and intensive care, due to the rareness and diversity of underlying disorders, wide age ranges, and the corresponding diversity of the age-specific normative values, if available at all. If such evidence is available, the design of knowledge-based CDSS implementing well-known criteria and rules is common. However, in cases of lacking evidence, knowledge acquisition for rule construction is crucial, which is highly time consuming and dependent on the knowledge and experience of the respective experts. Together with the increased amount of data available, research on the use of data-driven approaches such as in machine learning has been reinforced, also in the context of sepsis detection. Furthermore, such approaches might allow not only for detection of the disease at the time of onset but also for predicting the condition before recognizing symptoms in clinical routine. Over the last years, many different CDSS have been developed, introduced successfully, and evaluated in all kinds of settings and patient populations with highly varying results. In our work, we aim at providing a systematic overview of current systems and their characteristics, published in the context of computerized clinical decision-support for detection of SIRS and related conditions such as sepsis, severe sepsis, and septic shock.

Objectives

We present the results of a systematic literature review focusing on (1) the identification of computerized tools (*intervention*) for detection of SIRS, sepsis, severe sepsis, or septic shock in critically ill patients (*population*), (2) a classification of the identified approaches into predefined categories in a standardized format (*outcome*), and (3) the presentation of the reported performances compared with noncomputerized methods and other algorithms (*comparison*).

Related work

Early detection and prediction of SIRS and sepsis is a well-known research theme that already has been a topic of interest in other literature reviews. The latest publication in this context, published in 2019 by Islam et al, presents "(…) a meta-analysis of observational studies to quantify the performance of a machine learning model to predict sepsis."[23] The review was conducted systematically in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and yielded seven studies, which were included for comparison of performances. The reported accuracies, including sensitivities and specificities, were extracted and analyzed leading to the conclusion that machine learning algorithms had a better performance than traditional sepsis scoring systems.[23] Due to its systematic nature, the review provides results that are interesting, reproducible, and comprehensible. However, the search term might be too restrictive because first, it consisted of MeSH terms only, and, second, the use of synonyms was lacking. Furthermore, the initial search resulted in only 135 articles retrieved out of four electronic databases.[23] Hence, it might be worth conducting an extended literature search.

In contrast to the aforementioned review, another recently published literature review from 2018, performed by Warttig et al,[24] focuses not only on machine learning algorithms but also on any automated system capable of screening patient records in the context of sepsis. The authors performed a systematic review by searching in eight databases with a broad search term. However, the authors finally only included studies presenting a randomized controlled trial (RCT) to assess the systems' impact (1) on reduction of time to appropriate treatments and (2) on improvement of clinical outcomes such as mortality. Although the initial search resulted in 3,233 records, only three articles (including two conference abstracts) were included in the qualitative synthesis. The authors assessed all studies as low-quality RCTs.[24] Furthermore, it was not easy to perceive whether these articles are independent of each other. The authors conclude that "it is unclear what effect automated systems for monitoring sepsis have on any of the outcomes in this review."[24]

Another overview of current sepsis screening tools was published by Villegas and Moore in 2018.[25] The authors identified six screening tools, presented their performances and, thereby, assessed the current state of the art. However, the searches as well as the presentation of results were not performed systematically so that some relevant studies might be missing. For our research, this overview is not sufficient enough. However, the results from this overview were used as a source for retrieving relevant articles for our review.

In 2017, Despins published a related systematic review entitled "Automated detection of sepsis using electronic medical record data."[26] By searching PubMed between 2005 and 2015, the author was able to identify 13 relevant studies out of 222 initial search results. The used search term was restricted on one database (PubMed) and only included the search for electronic medical records/EHRs and sepsis. No other synonyms or phrases were used for initial search and no second reviewer was involved. However, in the synthesis and analysis, the author not only considered performance aspects but also other categories (e.g., patient sample setting and data used). The author concludes that an automated sepsis detection might be a possibility to enable early therapy but "(…) the evidence from the reviewed studies is inconclusive."[26] This

work is an appropriate basis for conducting our enhanced literature review, in which we do not restrict the search term on electronic medical records. Furthermore, we pursue an expanded search in additional databases.

For our systematic review, we are geared to the work presented by Makam et al in 2015.[27] The authors provide an enhanced, systematic review determining the diagnostic accuracy and the effectiveness of automated electronic sepsis alert systems. They searched various databases with an appropriate search term until June 2014 and retrieved 1,293 initial articles, resulting in eight eligible studies. The included studies were extensively analyzed and categorized according to their characteristics (e.g., setting, alert threshold, gold standard definition, alert system type, study quality). In the end, the authors reason that automated sepsis alerts may improve care processes but "(…) tend to have poor positive predictive value (PPV) and do not improve mortality or length of stay."[27] In our work, we aim at updating this review because much progress has been made in the last 5 years. Furthermore, we want to enhance this review by considering systems not only for adult care but also for neonatology and pediatrics.

Methods

Protocol
The systematic review was based on the PRISMA statement.[28]

Eligibility Criteria
All articles retrieved were assessed independently by two reviewers (AW, SM) according to the following inclusion criteria: (1) strong focus on SIRS, sepsis, severe sepsis, or septic shock, (2) detection or prediction by using an automated electronic system or algorithm, (3) evaluation of the diagnostic accuracy of the system or algorithm (sensitivity, specificity), and (4) processing of routine data as vital signs or routine laboratory values originating from the EHR or related sources.

We excluded articles that (1) focus on a specific underlying disease or on infections in general, (2) present explorative analyses and experiments with machine learning algorithms aiming at gathering new insights on valuable parameters rather than evaluating the diagnostic accuracy of an algorithm or a holistic system, (3a) only evaluate the impact of decision-support systems on mortality, sepsis management, or timings of therapy, (3b) do not report a sufficient evaluation methodology, including ground truth definition and results, and (4) examine biomarkers or related not routinely measured parameters only.

Information Sources and Search
We retrieved relevant articles through literature searching in five electronic bibliographic databases, including PubMed/MEDLINE, IEEE Xplore, Embase, Scopus, and ScienceDirect. There were no restrictions to language, publication type, state, and date. The search strategy was constructed according to the PICO framework, whereat *outcome* and *comparison* were not explicitly covered in the initial search term. The

term addresses computerized tools and systems for supporting the detection or prediction of patients suffering from SIRS, sepsis, severe sepsis, or septic shock. We decided to use four "AND"-concatenations: the first block represents the search for tools, systems, applications, platforms, or algorithms, the second emphasizes the computerized or electronic type of the aforementioned tools, the third defines the aspired functionality of the searched systems (e.g., detection, prediction, alerting), and the fourth covers the patient population. All search terms were restricted to the title, the abstract, and/or the keywords of the articles. When applicable, MeSH terms were used in addition to the standard search. In each block, we tried to cover as many synonyms as possible. Hence, the following search term was constructed for PubMed/MEDLINE:

(Decision Support Systems, Clinical[MeSH Terms] OR Decision Support Systems, Management[MeSH Terms] OR Expert Systems[MeSH Terms] OR system[Title/Abstract] OR tool[Title/Abstract] OR software[Title/Abstract] OR engine[Title/Abstract] OR approach[Title/Abstract] OR algorithm*[Title/Abstract] OR application[Title/Abstract] OR platform[Title/Abstract] OR method[Title] OR trial[Title/Abstract])
AND
(automated[Title/Abstract] OR technology assisted[Title/Abstract] OR Artificial Intelligence[MeSH Terms] OR computer*[Title/Abstract] OR computational[Title/Abstract] OR data mining [Title/Abstract] OR Data Mining [MeSH Terms] OR machine learning[Title/Abstract] OR Machine Learning [MeSH Terms] OR electronic*[Title])
AND
(Decision support[Title/Abstract] OR decision aid[Title/Abstract] OR Decision Making, Computer Assisted[MeSH Terms] OR Diagnosis, Computer-Assisted[MeSH Terms] OR detect*[Title/Abstract] OR predict*[Title/Abstract] OR warn*[Title/Abstract] OR alert*[Title/Abstract] OR recog*[Title/Abstract] OR screen*[Title/Abstract] OR monitor*[Title/Abstract] OR surveillance[Title/Abstract] OR assess*[Title/Abstract] OR diagnos*[Title/Abstract])
AND
(systemic inflammatory response syndrome[MeSH Terms] OR systemic inflammatory response syndrome[Title/Abstract] OR Shock, Septic[MeSH Terms] OR septic shock[Title] OR Sepsis[MeSH Terms] OR sepsis[Title/Abstract])

All queries for the other databases were constructed in alignment with this query. However, some adaptions were required to yield relevant results. For example, medical use cases are rare in the technical database IEEE Xplore. Hence, the corresponding search term consisted of the patient population and the disease block only. For Embase and Scopus, the sepsis term was restricted on title search only because sepsis indeed is very often mentioned as a prominent example but the original article focuses on another condition. Furthermore, in some databases, the length of the search term is restricted so we were not able to utilize all synonyms in all databases. The search terms are presented in ►**Supplementary Appendix A.1**.

Additional articles were identified by screening the reference lists and citations of all included studies (*backward*

searching). The relevant systematic reviews (see Chapter 1, Related Work) also were used for identifying further articles.

Study Selection
The study selection process is presented in a PRISMA flow diagram[28] (see ►Fig. 1). After performing the database searches, all results were exported to Citavi 5.7.1.0 and duplicates were removed. Subsequently, two authors (AW, SM) independently screened the titles and the abstracts for relevance and inclusion in the review by utilizing the eligibility criteria described in the Methods section. Afterwards, we retrieved full-texts and also evaluated them in accordance with the predefined criteria. The reasons for exclusion of the articles were documented (see ►Supplementary Appendix A.2). In any case of disagreement or discussion, a third author (TJ) gave advice. Backward searching by the same reviewers took place after all full-texts were assessed and the decisions on inclusion were made.

Data Extraction, Quality Assessment, and Measures
Following information was extracted from each included article: (1) general aspects including authors, title, publication year, and journal, (2) clinical settings of the study, e.g., ICU or emergency department (ED) and mono- or multi-centric, (3) the patient population including the age group, inclusion and exclusion criteria, and/or underlying diagnosis, (4) the focused condition as SIRS, sepsis, severe sepsis, and septic shock, (5) characteristics of the supporting system including the type of support (detection or predic-tion) and the underlying algorithm, (6) characteristics of data processed by the system or used within the evaluation, (7) evaluation methodology including description of ground truth and sample size, and (8) evaluation results including sensitivity, specificity, PPV, and negative predictive value (NPV). Two reviewers (AW, SM) extracted data from the included articles. Discrepancies between the reviewers were resolved by discussion and support of a third reviewer (TJ).

Synthesis of Results
We performed a qualitative assessment of the included articles. To give a holistic overview of current approaches, a table of characteristics was created for eligible articles (see ►Supplementary Appendix A.3). Data were summarized and articles were categorized in a standardized way. Thereby, the following categories and attributes were used:

- Clinical setting:
  - Ward: ICU, ED, non-ICU, all.
  - Special characteristics: age group (premature, neo-nates, pediatrics, adults; including age restrictions if applicable), inclusion and exclusion criteria for studies and evaluations.
  - Institutions: monocentric, multicentric.
- Support:
  - Type: detection (*detection at or after onset*), prediction (*detection before onset*).
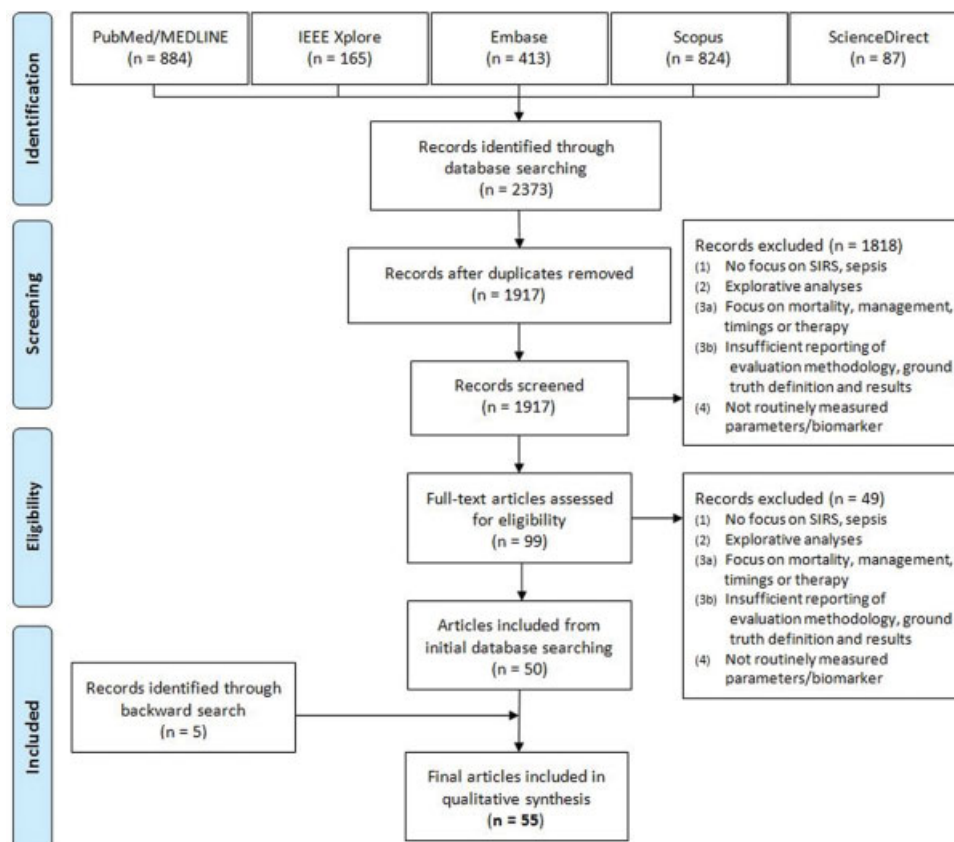 - Condition: SIRS, sepsis, severe sepsis, septic shock.

Fig. 1 PRISMA flow chart for study selection. PRISMA, preferred reporting items for systematic reviews and meta-analyses.

- System:
  - Name of the system
  - Underlying algorithm: threshold/rule/criteria-based, machine learning (including the specific algorithm).
  - Data/type of source system: own system/collected data or public database (e.g., Medical Information Mart for Intensive Care [MIMIC]).
- Evaluation:
  - Ground truth definitions: manual chart review, clinical experts, diagnosis coding, criteria (e.g., SOFA, SIRS), culture positivity, antibiotic administration.
  - Results: sample size, sensitivity, specificity, PPV, NPV.

Due to the highly heterogeneous trial conductions and ground truth definitions, we were unable to perform a meta-analysis. However, if applicable, we report the diagnostic accuracies of the systems by reporting sensitivity, specificity, PPV, and NPV, and qualitatively assess the evaluation methodologies and results.

Ethical Considerations

This manuscript does not contain research involving human subjects.

Results

Study Selection

An initial literature search in 2018 together with an updated search in January 2019 resulted in 2,373 articles (see ►Fig. 1). These were extracted from the databases PubMed/MEDLINE ($n = 884$), IEEE Xplore ($n = 165$), Embase ($n = 413$), Scopus ($n = 824$), and ScienceDirect ($n = 87$). After removing duplicates, 1,917 articles were screened by titles and abstracts for relevance in accordance with the eligibility criteria. The excluded articles are summarized together with the exclusion reason in ►Supplementary Appendix A.1. In addition, abstracts, short conference articles (less than two pages), editorials, reviews, and posters were removed. The full texts of 99 articles were retrieved and evaluated. Of these, a total of 50 articles were found eligible for inclusion. An additional backward search was performed leading to five more articles found to be eligible. Finally, 55 articles were included for the qualitative synthesis (see ►Supplementary Appendix A.3).

Study Characteristics

Clinical Settings

In total, 73% of the articles describe studies conducted in ICU, including pediatric ICU and neonatal ICUs (NICUs), or EDs. Non-ICUs are covered in 7% of the articles whereas 20% of the studies were not restricted to a specific ward. A similarly clear situation can be observed in the context of the age group of the patients included (see ►Fig. 2). More than 85% of all eligible articles included adults. A total of 51% of all articles comprise studies on adults only, which we defined as patients ≥18 years. Other explicit age restrictions covered are ≥14 years (9%), ≥15 years (4%), and ≥16 years (2%). Some

AGE GROUPS COVERED IN ARTICLES
CATEGORY: CLINICAL SETTING

Legend:
- PREMATURE
- NEONATES
- PEDIATRICS
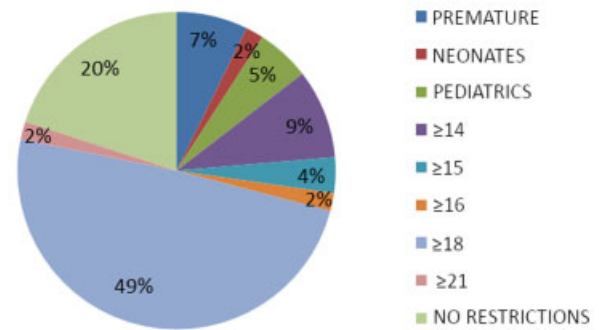- ≥14
- ≥15
- ≥16
- ≥18
- ≥21
- NO RESTRICTIONS

Fig. 2 Distribution and frequency of age groups covered in articles (category: clinical setting).

articles focus on specific cohorts as very low birth weight (VLBW) and/or premature newborns of ≤28 or ≤32 weeks of gestational age (7%), neonates (2%), and pediatric patients (5%). However, the latter often is not defined explicitly by an age restriction so that possibly these articles also include patients of the categories ≥14, ≥15, and ≥16 years.

During research for this review, it became obvious that there are established algorithms for adult patients and neonates but there is a lack of systems supporting pediatric patients. There is no established commercially available system that was developed for this population particularly. There are only four studies that were performed explicitly on pediatric patients. None of these studies gave any detailed information about the algorithm used. Two of these focused on patients <18 years old,[29,30] one included "pediatric patients"[31] and one included mainly patients <18 years old but also 17% of patients ≥18 years old.[32] Cruz et al[30] describe a "best practice alert" in the ED that targeted septic shock using SIRS criteria in a single-center study. The algorithm was searching for tachycardia, as defined by the age-dependent 2005 International Pediatric Sepsis Consensus Conference (IPSCC) SIRS criteria, but adjusted by a temperature correction, which was not evaluated separately. Patients were enrolled if the alert triggered or if there was the diagnosis of shock during the further hospital stay. The further procedure was determined by the attending nurse.

Sepanski et al[29] describe an algorithm that targets the age-dependent IPSCC SIRS criteria to screen for severe sepsis. This study included a refinement of these criteria to improve specificity that took a rather large dataset of 143,603 cases into account. These patients visited the ED of one hospital. The refinement included a temperature correction of heart rate and respiratory rate and an adjustment of the threshold values of most SIRS criteria due to the data acquired of this patient group. The developed screening tool was implemented in another hospital. This Screening Tool Refinement Group consisted of 7,402 patients admitted to the ED during the appointed duration of the study. Ground truth for severe sepsis was a systemic chart review for patients with discharge diagnosis of "disseminated or localized infection/associated"[29] condition. According to the stated purpose, the sensitivity of the screening tool with the refined SIRS criteria was equal to the results of the IPSCC

criteria (97.4%), but the specificity and PPV were higher with 99.5 and 48.7% compared with 97.1 and 14.6%, respectively.

There is a proof-of-concept study published by Wulff et al[31] that used a rule-based algorithm based on the official 2005 IPSCC SIRS criteria for SIRS and sepsis detection in ICU patients. Due to this being a pilot study, there was just the small number of 16 patients enrolled and it was performed at a single hospital. The ground truth was defined by a blinded chart review done by two reviewers, who also applied the SIRS criteria for the detection of SIRS and sepsis in the included patients. The system had a high sensitivity of 100% and a specificity of 94%.

The latest publication of these four is a pilot-study done by Lloyd et al[32] who evaluated the electronic version of an already established manual sepsis screening tool in a pediatric ED. This tool is described as a modified version of a sepsis trigger and identification tool from the Pediatric Septic Shock Collaborative of the American Academy of Pediatrics. During an 8-week-period, they identified 29 sepsis patients recognized by both the system and the manual screening tool. However, the electronic version was significantly faster and, on average, flagged the patients 68 minutes earlier.

In total, 42% of the studies included all admitted patients, 44% included patients with a diagnosis related to sepsis, suspected infection on admission, or with an acquired blood culture, and 9% included only patients with a positive alert. Eleven percent recruited a very specific preselected population as preterm VLBW babies, patients with burns or after cardiothoracic surgery.

In terms of the number of participating institutions, most studies took place at a single site only (78%). However, there are also some studies that encompass a huge number of patients from two (7%) or more institutions (15%).

Support

From a technical perspective, we decided to distinguish between *detection* and *prediction* of diseases. The former is defined as any system able to detect sepsis or a related condition at the time of onset or after onset. In contrast, the latter represents algorithms able to forecast any condition onset. Detection only is covered in 62% and prediction only is focused in 33% of the included articles. Three articles (5%) present algorithms covering both detection and prediction. In the context of the clinical domain, a differentiation between SIRS, sepsis, severe sepsis, and septic shock is appropriate. Here, 71% of the studies focus on the detection or prediction of sepsis (including subgroups of late-onset sepsis or neonatal sepsis), 5% focus on SIRS, 20% have been trying to identify severe sepsis, and 18% septic shock. Eleven percent present algorithms for the detection or prediction of more than one condition, in which the combination of *severe sepsis and septic shock* is the most common one. Two articles present a combination of sepsis, severe sepsis, and septic shock, and one article focuses on the combination of SIRS and sepsis. SIRS only is covered in 4%, sepsis in 67%, severe sepsis in 11%, septic shock in 9%, and any combinations in 11% of the included articles.

System

The documentation of the system's characteristics and the underlying algorithmic approach were defined as important eligibility criteria. In terms of the underlying algorithm, systems following a traditional knowledge-driven and, a therefore, threshold-, criteria-, or rule-based approach were distinguished from tools implementing a data-driven, machine learning concept. For the latter, a high variety of specific
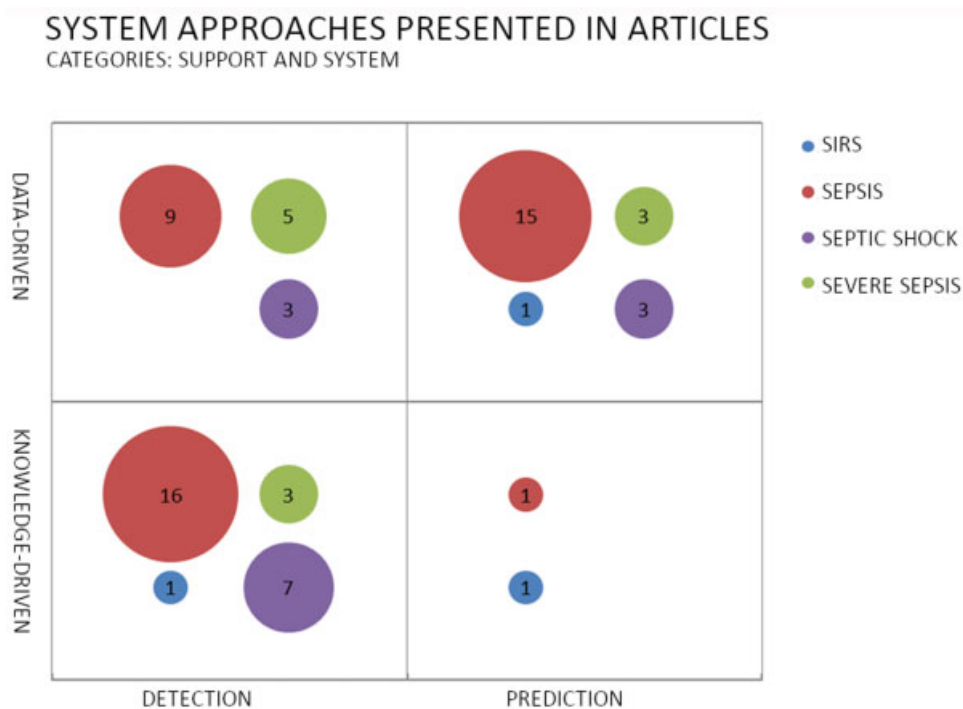

Fig. 3 Distribution and frequency of system approaches presented in articles (categories: support, system).

algorithms such as random forests, neural networks, decision trees, support vector machines, and others were used in the included articles. Fifty-six percent of all articles deal with machine learning approaches and 44% implement knowledge-based concepts. In addition, some authors present a comparison of the core algorithm to other approaches, including both machine learning models and sepsis scoring systems and respective published criteria. All information can be retrieved from the ►Supplementary Appendix A.3.

To achieve a better understanding of the relationship between the support type (detection/prediction and SIRS/sepsis/severe sepsis/septic shock) and the underlying algorithmic approach (knowledge-based/data-driven), ►Fig. 3 presents the distribution of frequencies for each combination. The most common combinations discovered are knowledge-based detection of sepsis and data-driven prediction of sepsis. Threshold-/criteria-/rule-based approaches for prediction of disease seem not to be very common, whereas machine learning approaches are almost equally used for both detection and prediction tasks.

Another interesting characteristic of the systems is the used data source. Only some systems already provide a seamless integration into the EHR or any other routine system. However, none of the articles provides a detailed description of the system architecture including the interfaces and dashboards. Furthermore, no other articles other than that by Wulff et al[31] describe possibilities to easily implement the system into other institutions such as by using interoperability features for data representation and sharing of the algorithm. Only one further article by Henry et al,[33] which was excluded due to an insufficient evaluation, explicitly considered FHIR (Fast Healthcare Interoperability Resources) as a possible standard ("A FHIR-enabled streaming sepsis prediction system for ICUs"). In most of the articles, the systems were tested on datasets extracted from a primary source system (e.g., EHR, PDMS or bedside monitor systems) and entered into the system's algorithm for the purpose of the study. In 20% of the articles, the authors reused data from a publicly accessible database called MIMIC[34] rather than generating own datasets from local systems.

Evaluation

All included articles present a sufficient evaluation methodology together with reporting of results, including at least sensitivity/specificity and/or PPV/NPV. However, amongst others, the trials are highly heterogeneous according to the sample size, the ground truth definitions, and the inclusion or exclusion of patients. As this does not allow performing an enhanced meta-analysis, we decided to provide a qualitative analysis of the evaluations.

In terms of the used ground truth, the included articles can be divided into four major groups (see ►Fig. 4). Thirty-eight percent of the studies used manual chart reviews, involved clinical experts, or assessed according to well-known criteria (e.g., sepsis definition by Goldstein et al[4]; *group 1: criteria*). A slightly smaller group (29%) defines positive cases as having a coded diagnosis of sepsis, severe sepsis, or septic shock (e.g., International Classification of Diseases [ICD] coding, diagnosis-related group [DRG] coding; *group 2: coding*). In 20% of all included articles, various techniques were combined (*group 3: mixed*). Here, the combination of criteria and coding is the most prominent (54%), followed by criteria and blood culture (18%) as well as blood culture and antibiotic administration (18%, Mitchell et al,[35] Wang et al,[36] and Nemati et al[37]). One article uses a special method for distinguishing septic patients consisting of ICU transferral, activation of rapid response teams, and coding.[38] The fourth and smallest group of articles defines the condition onset only if the patient has a positive blood culture (13%, Gur et al,[39] Mani et al,[40] Van Steenkiste et al,[41] Guillen et al,[42] McCoy and Das,[43] Mithal et al,[44,45] and Wang et al[46]; *group 4: blood culture*).

Furthermore, the included articles differ in terms of the sample size, which is the number of patients, visits, screens, or admissions included in the evaluation. The sample size ranged from a pilot study with eight patients to a multicentric study with 2,759,529 patients from 49 institutions. The latter is the only evaluation with a sample size of more than 1 million. As illustrated in ►Fig. 5, most of the articles present evaluations with a sample size between 10,001 and 100,000.
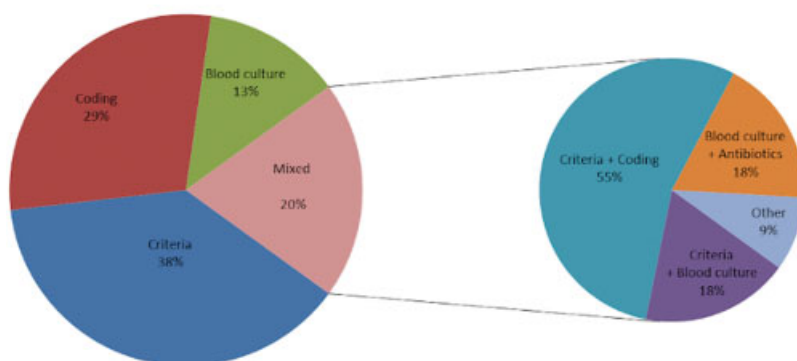

Fig. 4 Distribution and frequency of ground truth definitions used in articles (category: evaluation).

SAMPLE SIZES USED IN ARTICLES
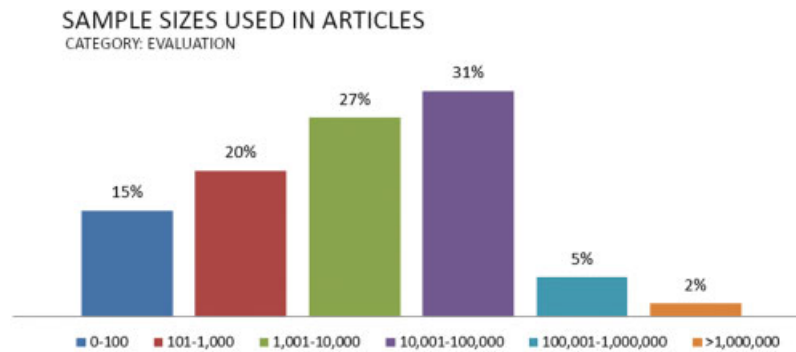CATEGORY: EVALUATION

Fig. 5 Distribution and frequency of sample sizes used in articles (category: evaluation).

The evaluation results consisting of sensitivity, specificity, PPV, and NPV can be found in ►Supplementary Appendix A.3. For articles testing more than one algorithm, we decided to report the results for the best algorithm. For articles providing results for different cut-off values, fixed sensitivities or specificities, and time windows, if possible, we reported all results in our table. Altogether, we identified five articles presenting a sensitivity and a specificity of more than 90%: Alsolamy et al,[47] Sepanski et al,[29] Kam and Kim,[48] Bansal et al,[49] and Wulff et al.[31] These articles are alike in that they all present single-center studies, used data from the local EHR instead of publicly available datasets, and developed detection rather than prediction algorithms. Wulff et al[31] focused on pediatric SIRS detection, Sepanski et al[29] yielded at pediatric severe sepsis detection, Kam and Kim[48] presented sepsis detection for adults and Bansal et al[49] the same without age restrictions, Alsolamy et al[47] developed an algorithm for severe sepsis and septic shock detection in patients older than 14 years. The study of Kam and Kim[48] is the only one presenting a data-driven approach (using neural networks and long short-term memory) and using coding as ground truth. All other articles present a threshold-/criteria-/rule-based algorithm together with a manual chart review as ground truth. The sample sizes vary between 16 and 49,838 cases included.

Discussion

Summary of Evidence
Recently, the number of publications addressing the design of systems supporting clinicians in detection and prediction of SIRS and related conditions has increased rapidly. Although most of the studies aim toward reaching the same ultimate goal of improving the patient's chances in the end, the articles' approaches to provide evidence vary. Lately, there seems to be a rather direct approach to prove the effect of a system's implementation on either reduction of length of stay or clinical outcomes as mortality, or success of sepsis response teams, sepsis bundle adherence, or therapies.[24] More conservative approaches first aim at assessing the diagnostic accuracy of a supporting system before targeting on demonstration of effects on clinical outcomes. As the aforementioned article group mostly does not focus on the algorithms itself, in our systematic review, we decided to concentrate on the latter.

Furthermore, in the end, any improvement for the single patient only is possible if the algorithm is able to accurately detect the disease, resulting in very high sensitivity and specificity needed. Only then, it may be integrated in the daily clinical routine, offering a realistic chance to prove the effect of CDSS to improve treatments and outcomes.

Even before creating a system stands the question of using published evidence together with expert knowledge to design a knowledge-driven, often rule-based, system, or applying data-driven, machine learning algorithms. Sorting the included articles of our review according to their publication date, and correlating it with the algorithmic approach used, a trend starting in 2012 becomes visible (see ►Fig. 6). While there is no clear trend recognizable related to the quantitative number of publications, the portion of data-driven approaches is slightly increasing.

In terms of knowledge-based systems, our systematic review demonstrates the challenge of choosing the criteria on which the algorithm can be based upon. Hence, a high variety of approaches is used, which complicates any comparisons of outcomes between the articles.

Although described as highly unspecific, many authors used the SIRS criteria for sepsis detection, so that it was not surprising to discover that many publications report alert fatigue due to a low specificity. Publications using the Sepsis-3 guideline often adapt the SOFA score focusing on signs for organ failure and, therefore, reaching higher specificity, which is mostly due to the further progressed course of sepsis. Some of the studies presented in this review tried to improve sepsis recognition by defining new criteria and using new markers. Sepanski et al[29] were making huge efforts to empirically identify new vital sign thresholds for the pediatric SIRS criteria using a "vital sign standardization group" consisting of 143,603 patients. Cruz et al[30] simply used a temperature correction on the pediatric SIRS criteria to take the impact of elevated core temperature on both heart rate and respiratory rate into account. Lloyd et al[32] utilized adapted parameters from the Pediatric Septic Shock Collaborative of the American Academy of Pediatrics for their electronic version of the manual screening tool. Mithal et al[45] and Gur et al[39] were using RALIS, which creates a RALIS score based upon the heart rate, respiratory rate, temperature, desaturation, and bradycardia for gestational age in neonates. The precise values and details about the
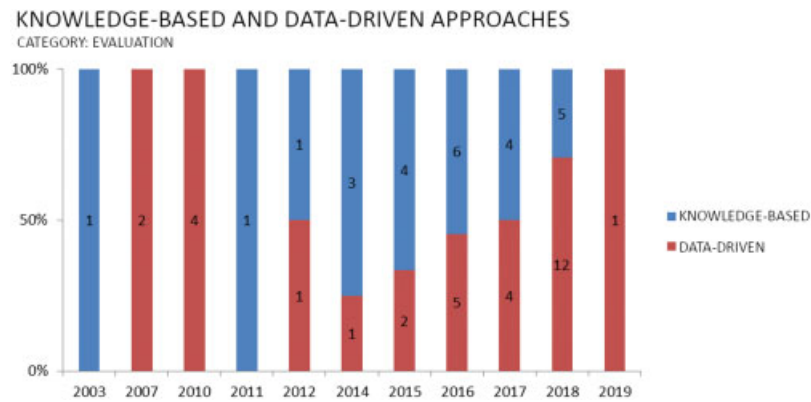
Fig. 6 Distribution and frequency of knowledge-based and data-driven approaches per publication year (category: system).

algorithm could not be found, possibly due to its commercial value as an already patented product.

The variety of criteria, which potentially can be used to detect SIRS or sepsis, the often low evidence, and otherwise great efforts it takes to reach reliable results, especially in pediatrics and neonatology, might be reasons for the increased use of data-driven approaches. Here, the criteria and classification rules are constructed by the algorithm itself, which resulted in a high number of attributes used for reasoning. However, this also means that a high input of data has to be available for training and implementing the algorithms. This is particularly required for the prediction of sepsis, severe sepsis, or septic shock. For example, Mani et al[40] compared nine different machine learning algorithms and used 781 temporal and 30 nontemporal variables to predict neonatal sepsis. Khojandi et al[50] used 57 features of categorical and continuous variables for sepsis prediction, whereas Lin et al[51] describe the use of 35 static and 43 dynamic variables to predict septic shock. To predict sepsis in adults, Nemati et al[37] used 65 features, Saqib et al[52] 47 features, and Wang et al[36] utilized 43 different variables. Henry et al[53] created their own targeted real-time early warning score employing 54 potential features from the EHR to predict septic shock. Brown et al[54] used modified SIRS criteria and additionally 75 parameters for detection of severe sepsis and septic shock. Gunnarsdottir et al[55] tested a different approach by analyzing heart rate variability to reach a sepsis decision. In the data-driven group, there also is a commercial product, called InSight, which promises an earlier sepsis prediction, improvement in outcome, and reduction of costs. According to the advertisement, it has also been validated for a highly diverse pediatric population. This system creates its own score and is used by Calvert et al,[56] Desautels et al[57] and lately Mao et al.[58] Overall, in several publications, the exact criteria used are not even reported. This represents a major obstacle for comparison, comprehension, and quality of results.

The use of CDSS in a clinical routine highly depends on its diagnostic accuracy, which should be evaluated under settings as realistic as possible. One important aspect is the definition of an accurate and realistic ground truth to evaluate the system's performance. As our review shows, here lies one of the largest discrepancies between the studies: 29% of the articles used a general manual chart review, and only 7% performed a blinded manual chart review by more than one reviewer. This, of course, is a very time-consuming but solid method.[29,31,39,59] Thirty-six percent of the articles defined their ground truth by ICD coding for sepsis or sepsis-related diagnoses. This procedure allows including a much larger population but it relies entirely on the assumption that every patient received the correct diagnosis, and any signs of sepsis have not been missed before. Often, the coded diagnosis does not reflect the condition of the patient to the relevant time in clinical routine. Therefore, this procedure can be assessed as inferior to a manual chart review. In 20% of the studies, the ground truth for sepsis is defined as a positive culture or a culture taken and antibiotic administered which fulfills part of the sepsis definition but cannot be used to detect SIRS and has a relevant risk of missing septic patients. Especially in children and neonates, a positive culture just occasionally occurs even though they definitely suffer from sepsis clinically.[60,61] Even in adult patients, a positive blood culture can be obtained in only 50% of patients with blood stream infection.[62] If antibiotic treatment was administered before blood culture acquisition, the sensitivity directly drops with the risk of missing sepsis episodes.[63] The reliability of blood cultures in general depends on many factors such as the timing, sample volume, amount of collected samples, interpretation of positive results, and the rate of contaminations.[64] Based on these considerations, a positive blood culture as singular parameter for sepsis seems not to be recommendable. Fifteen percent of the articles utilized clinical assessment with clinical experts to determine whether a patient had sepsis. Often, these experts were not enumerated in more detail. Other authors used a combination of the methods mentioned above, which seems to be an accurate approach.

When using supporting systems in clinical routine settings, the system should be able to perform accurately not only for a very specific patient population but also, in the best case, for all patients admitted to this setting. However, the systems presented in the articles are mostly designed and tested on a very specific and preselected patient population. With our review, we uncovered the wide variety of inclusion and exclusion criteria that has been used. Examples are Schuh[65] who included only patients after cardiothoracic surgery and Harrison et al[66]

who focused on detection of severe sepsis in ICU but excluded patients with ICU-acquired sepsis. For detection and prediction of sepsis, Desautels et al[57] excluded patients with an onset of sepsis during the first 7 hours after admission, and Mao et al[58] did so as well. Mithal et al[44] were targeting late-onset sepsis and included only premature newborns under 28 weeks of gestational age with a complete dataset for the first 28 days. Patients with early-onset sepsis, transferral or death during these first 28 days of life, congenital syndromes or multiple anomalies, or with a need for high-frequency oscillation were excluded. Thus, many patients that will be admitted to NICU would not qualify for the study and, consequently, independent from the study results, the algorithm may not be useful for the majority of patients in the NICU. This is particularly misleading as these patients even have a higher risk for late-onset sepsis than the test cohort. Rech et al[59] included patients in their burn-ICU with at least a total burn surface area of >10% and/or at least moderate inhalation injury, who have a very high probability to develop at least SIRS. A total of 41% of the publications used the admission or discharge diagnosis, or patients with proven or suspected infection, for inclusion in the study, and 7% included only patients with a positive alert; Nelson et al's[67] study was the only study of this group to present a countercheck of all ED visits for 1 exemplary week and, therefore, a search for false negatives. Khojandi et al[50] explain that they only admitted patients with pneumonia to their study to have a guaranteed infection and possibility of progress to sepsis. These restrictions in recruitment are very likely to have biased results in clinical routine settings.

In contrast to these heterogeneous approaches to define the target group, the situation is more even when looking at the age of the included patients. Our review revealed that more than 50% of the articles concentrate on adult patients. Only 15% of the studies recruited patients over 14, 15, or 16 years while making no explanation on why the cut-off was made at this age. Furthermore, in these as well as in all studies without age restriction, it was not mentioned whether these patients were treated as adult patients with respect to their normative ranges of vital signs, laboratory values, and others. Only the four articles presented in the results section explicitly covered the pediatric group with both specific criteria and approaches differing depending on the age.

Another great disparity between the presented studies is the aim of their algorithm; 62% focused on detection whereas 31% introduced a prediction system, and 5% aimed at both detection and prediction.[57–59] And even in these groups, there are differences of focus as some try to find earlier alterations as in SIRS, whereas others concentrate on detection of the high-risk patients with severe sepsis or septic shock. Less severe symptoms, as in the beginning of SIRS, are often less specific and, therefore, the specificity of such a system will naturally be lower as in systems that can rely on parameters for organ dysfunction as defined for severe sepsis.

When designing CDSS for future use in clinical routines, it seems to be beneficial to use real data in the evaluation phase already. Indeed, our review supported this assumption by showing that 80% of the articles generated own datasets from local systems. The other 20% of the studies rely on the use of publicly accessible databases such as MIMIC, which seems to be an easy way to test algorithms. However, such datasets often come with a high data quality not available in unprepared routine data. Consequently, algorithms tested with such datasets yield outcomes, which are often not attainable in clinical routine settings. Even when using real datasets originating from EHRs or similar records, data quality should be considered as one very important aspect affecting and, in the end, determine the quality of results. In several articles, a full dataset is required or continuous data for a determined period of time, in most cases even several hours. Such datasets are often only available if explicitly recorded in the context of a clinical trial. Furthermore, in most publications, vital sign acquisition and validation was not described in detail. Most authors just extracted any data from the EHR or from a dataset; some describe manual vital sign acquisition if automated vital signs were not available. Neither publication mentions regular validation of data in a manual or automated way to take artifacts or errors in deduction into consideration, even though some refer to previous work that might cover some form of validation. Most algorithms have integrated thresholds to exclude implausible data, which were not explained in greater detail.

Some of the articles, as in Nemati et al,[37] Mitchell et al,[35] and Mao et al,[58] present a dual approach by testing the algorithm on both own datasets and MIMIC. This approach is advantageous because reaching good results on both datasets demonstrates that the algorithm is not overly fitted toward a local patient cohort. The same can be reached when conducting a multicentric study as presented in 22% of the included articles.

As discussed before, the ground truth definition is highly important when evaluating the results of the algorithms. In data-driven approaches, this labeling often is already available within the dataset used for training the algorithm. Consequently, one has to bear in mind that the algorithm always learns the cases, in which the label shows a specific value. For example, when using ICD coding, this means that the algorithm will learn when a sepsis code arises, based on all other attributes. As explained above, these are rarely the real sepsis events in clinical routine settings. However, a huge amount of —not necessarily but preferably labeled—dataset is needed for advanced machine learning, which is often not accomplishable by manual chart review or clinical assessments. Only a few articles present such an approach, as in Mani et al,[40] Nachimuthu and Haug,[68] and Brown et al.[54]

When looking at the good results reported for detection of adult sepsis by knowledge-based systems, together with the resource-consuming training of data-driven algorithms, the use of machine learning for detection purposes definitely is questionable. Furthermore, the so-called *black box* approach still is not seen as trustable in clinical routine settings, especially when there are explainable approaches with same or even better results. However, machine learning approaches are highly valuable when trying to predict patients' conditions as well as when uncovering yet unknown patterns and correlations between parameters. The former already could be observed in our review as most of the prediction tasks were performed by data-driven

approaches. The latter was not covered within this review as we decided to exclude any articles presenting machine learning-based explorative analyses. However, a set of these articles can be found in ►Supplementary Appendix A.2.

Up to now, the prediction and detection of pediatric SIRS is underrepresented in research. Since the introduction of the Sepsis-3 guidelines, the definition of SIRS or rather its clinical relevance seems to have faded. Moreover, recent studies evaluating different sepsis scores with regard to mortality reported that SIRS criteria have lower accuracy compared with SOFA, qSOFA, or MEWS (Modified Early Warning Score).[69–71] Consequently, evaluating SIRS criteria for prediction of mortality seems like the wrong approach. However, the big advantage of SIRS criteria for recognizing deterioration on a very early point in time has been overlooked lately. Of course, the SOFA score provides more accurate results in terms of representing the risk of mortality but it is not applicable as early as SIRS criteria. An early recognition of SIRS and sepsis can often trigger an early goal-directed therapy resulting in a prevention of situations, until which the other scores would not even be applicable.

Strengths and Limitations

Recently, a lot of CDSS have been developed and implemented in the context of SIRS, sepsis, severe sepsis, and septic shock. In our systematic review, we successfully retrieved and qualitatively assessed a wide scope of current studies. Thereby, we were able to carve out the differences between various approaches and, hence, presented the high variety of studies available. One of the major strengths of this review is its broad search approach including querying of five databases and using broad search terms without restricting the patient population on adults as done before by existing reviews. Hence, we updated and enriched the overview on the state of the art for CDSS in detection of SIRS and related conditions, available thitherto. By our approach, we successfully discovered a lack of research on SIRS detection and prediction in the pediatric population, which opens up major possibilities for future high-impact research.

Our systematic review has several limitations. To get an adequate overview, we included a large number of studies, which made it impossible to statistically analyze the study outcomes as required in a meta-analysis. As presented, the included articles followed very different study designs, sample sizes, evaluation approaches, and recruiting approaches, which also lead to a high variety of study quality. Although meeting our inclusion criteria, it should be kept in mind that several study designs are predisposed for biased results, as outlined in our summary of evidence.

Conclusion

Current research on CDSS for SIRS, sepsis, severe sepsis, and septic shock is characterized by a high variety of approaches. Several designs for either detecting or predicting the aforementioned conditions exist, sometimes even mixed up, evaluated by using various ground truth definitions, sample

sizes, and patient cohorts. Although the evaluations of knowledge-based solutions are not unambiguously bad, a clear trend toward the use of data-driven approaches is recognizable. On the contrary, a lack of research in pediatric patients and SIRS can be observed.

As a result of this review, in our future work, we will concentrate on supporting SIRS decision making by CDSS. We will attach great importance to the realistic evaluation of the diagnostic accuracy in routine clinical settings. Later on, data-driven methods should be evaluated on both institution-specific and cross-institutional settings to gather new insights. Therefore, after accurate evaluations, the rule-based detection algorithm might be usable to generate datasets based on retrospective data, resulting in a great training base with reliable SIRS and sepsis labels. In our review, we observed that there is a lack of interfaces to local EHRs and PDMS, live systems, interoperability features to share data, results, and even the system itself and its seamless integration into the clinician's workflow. As all of these factors strongly influence the success of implementing a CDSS in clinical routine, we will concentrate on these aspects in our future work on CDSS.

Authors' Contributions

AW and SM were equally responsible for conducting the overall literature review, including the design the methodological search strategy, and drafting the manuscript. Both AW and SM performed study selection and data extraction. TJ provided clinical expertise for constructing the initial search term and acted as a third reviewer in any case of disagreement. MM gave valuable advice for designing the methodological approach, revised the manuscript critically, supported in provisioning of full-texts and translations, and gave further subject-specific guidance on medical and technical topics. All authors read and approved the final manuscript.

References

1 Goto M, Al-Hasan MN. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. Clin Microbiol Infect 2013;19(06):501–509

2 Dellinger RP, Levy MM, Rhodes A, et al; Surviving Sepsis Campaign Guidelines Committee including The Pediatric Subgroup. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med 2013;39 (02):165–228

3 Chen Q-H, Shao J, Liu W-L, et al. Predictive accuracy of Sepsis-3 definitions for mortality among adult critically ill patients with suspected infection. Chin Med J (Engl) 2019;132(10):1147–1153

4 Goldstein B, Giroir B, Randolph A; International Consensus Conference on Pediatric Sepsis. International Pediatric Sepsis Consensus Conference: definitions for sepsis and organ dysfunction in pediatrics. Pediatr Crit Care Med 2005;6(01):2–8

5 Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med 2017;43(03):304–377

6 Cinel I, Dellinger RP. Advances in pathogenesis and management of sepsis. Curr Opin Infect Dis 2007;20(04):345–352

7 Kellogg JA, Ferrentino FL, Goodstein MH, Liss J, Shapiro SL, Bankert DA. Frequency of low level bacteremia in infants from birth to two months of age. Pediatr Infect Dis J 1997;16(04):381–385

8 Stranieri I, Kanunfre KA, Rodrigues JC, et al. Assessment and comparison of bacterial load levels determined by quantitative amplifications in blood culture-positive and negative neonatal sepsis. Rev Inst Med Trop São Paulo 2018;60:e61

9 Clapp DW. Developmental regulation of the immune system. Semin Perinatol 2006;30(02):69–72

10 Boehne M, Sasse M, Karch A, et al. Systemic inflammatory response syndrome after pediatric congenital heart surgery: incidence, risk factors, and clinical outcome. J Card Surg 2017;32(02):116–125

11 Schlapbach LJ, Straney L, Alexander J, et al; ANZICS Paediatric Study Group. Mortality related to invasive infections, sepsis, and septic shock in critically ill children in Australia and New Zealand, 2002–13: a multicentre retrospective cohort study. Lancet Infect Dis 2015;15(01):46–54

12 Wolfler A, Silvani P, Musicco M, Antonelli M, Salvo I; Italian Pediatric Sepsis Study (SISPe) group. Incidence of and mortality due to sepsis, severe sepsis and septic shock in Italian pediatric intensive care units: a prospective national survey. Intensive Care Med 2008;34(09):1690–1697

13 Han YY, Carcillo JA, Dragotta MA, et al. Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. Pediatrics 2003;112(04):793–799

14 Oliveira CF, Nogueira de Sá FR, Oliveira DSF, et al. Time- and fluid-sensitive resuscitation for hemodynamic support of children in septic shock: barriers to the implementation of the American College of Critical Care Medicine/Pediatric Advanced Life Support Guidelines in a pediatric intensive care unit in a developing world. Pediatr Emerg Care 2008;24(12):810–815

15 Levy MM, Fink MP, Marshall JC, et al; SCCM/ESICM/ACCP/ATS/SIS. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med 2003;31(04):1250–1256

16 Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22(07):707–710

17 Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315(08):801–810

18 Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the Sepsis-3 definitions in critically ill children. JAMA Pediatr 2017;171(10):e172352

19 Kwon R. Calculated decisions: pediatric early warning score (PEWS). Emerg Med Pract 2017(Suppl 2017B):12–13

20 Garcia PCR, Eulmesekian P, Branco RG, et al. External validation of the paediatric logistic organ dysfunction score. Intensive Care Med 2010;36(01):116–122

21 Wynn JL, Polin RA. Progress in the management of neonatal sepsis: the importance of a consensus definition. Pediatr Res 2018;83(1-1):13–15

22 Aakre CA, Kitson JE, Li M, Herasevich V. Iterative user interface design for automated sequential organ failure assessment score calculator in sepsis detection. JMIR Human Factors 2017;4(02):e14

23 Islam MM, Nasrin T, Walther BA, Wu C-C, Yang H-C, Li Y-C. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019;170:1–9

24 Warttig S, Alderson P, Evans DJW, Lewis SR, Kourbeti IS, Smith AF. Automated monitoring compared to standard care for the early detection of sepsis in critically ill patients. Cochrane Database Syst Rev 2018;6(06):CD012404

25 Villegas N, Moore LJ. Sepsis screening: current evidence and available tools. Surg Infect (Larchmt) 2018;19(02):126–130

26 Despins LA. Automated detection of sepsis using electronic medical record data: a systematic review. J Healthc Qual 2017;39(06):322–333

27 Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. J Hosp Med 2015;10(06):396–402

28 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6(07):e1000100

29 Sepanski RJ, Godambe SA, Mangum CD, Bovat CS, Zaritsky AL, Shah SH. Designing a pediatric severe sepsis screening tool. Front Pediatr 2014;2:56

30 Cruz AT, Williams EA, Graf JM, et al. Test characteristics of an automated age- and temperature-adjusted tachycardia alert in pediatric septic shock. Pediatr Emerg Care 2012;28(09):889–894

31 Wulff A, Haarbrandt B, Tute E, Marschollek M, Beerbaum P, Jack T. An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR. Artif Intell Med 2018;89:10–23

32 Lloyd JK, Ahrens EA, Clark D, Dachenhaus T, Nuss KE. Automating a manual sepsis screening tool in a pediatric emergency department. Appl Clin Inform 2018;9(04):803–808

33 Henry JR, Lynch D, Mals J, et al. A FHIR-enabled streaming sepsis prediction system for ICUs. Conf Proc IEEE Eng Med Biol Soc 2018;2018:4093–4096

34 Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035

35 Mitchell S, Schinkel K, Song Y, et al. Optimization of sepsis risk assessment for ward patients. Paper presented at: IEEE Systems and Information Engineering Design Symposium (SIEDS); April 29, 2016; University of Virginia, Charlottesville, Virginia, United States

36 Wang R, Sun C, Schroeder P, Ameko M, Moore C, Barnes L. Predictive models of sepsis in adult ICU patients. Paper presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 4, 2018; New York City, NY, United States

37 Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018;46(04):547–553

38 Umscheid CA, Betesh J, VanZandbergen C, et al. Development, implementation, and impact of an automated early warning and response system for sepsis. J Hosp Med 2015;10(01):26–31

39 Gur I, Markel G, Nave Y, Vainshtein I, Eisenkraft A, Riskin A. A mathematical algorithm for detection of late-onset sepsis in very-low birth weight infants: a preliminary diagnostic test evaluation. Indian Pediatr 2014;51(08):647–650

40 Mani S, Ozdas A, Aliferis C, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J Am Med Inform Assoc 2014;21(02):326–336

41 van Steenkiste T, Ruyssinck J, de Baets L, et al. Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. Artif Intell Med 2019;97:38–43

42 Guillén J, Liu J, Furr M, et al. Predictive models for severe sepsis in adult ICU patients. Paper presented at: Systems and Information Engineering Design Symposium; April 24, 2015; Charlottesville, Virginia, United States

43 McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017;6(02):e000158

44 Mithal LB, Yogev R, Palac H, Gur I, Mestan KK. Computerized vital signs analysis and late onset infections in extremely low gestational age infants. J Perinat Med 2016;44(05):491–497

45 Mithal LB, Yogev R, Palac HL, Kaminsky D, Gur I, Mestan KK. Vital signs analysis algorithm detects inflammatory response in premature infants with late onset sepsis and necrotizing enterocolitis. Early Hum Dev 2018;117:83–89

46 Wang S-L, Wu F, Wang B-H. Prediction of severe sepsis using SVM model. Adv Exp Med Biol 2010;680:75–81

47 Alsolamy S, Al Salamah M, Al Thagafi M, et al. Diagnostic accuracy of a screening electronic alert tool for severe sepsis and septic shock in the emergency department. BMC Med Inform Decis Mak 2014;14(01):105

48 Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017; 89:248–255

49 Bansal V, Festić E, Mangi MA, et al. Early machine-human interface around sepsis severity identification: from diagnosis to improved management? Acta Med Acad 2018;47(01):27–38

50 Khojandi A, Tansakul V, Li X, Koszalinski RS, Paiva W. Prediction of sepsis and in-hospital mortality using electronic health records. Methods Inf Med 2018;57(04):185–193

51 Lin C, Zhangy Y, Ivy J, et al. Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM. Paper presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 4, 2018; New York, NY, United States

52 Saqib M, Sha Y, Wang MD. Early prediction of sepsis in EMR records using traditional ML techniques and deep learning LSTM networks. Conf Proc IEEE Eng Med Biol Soc 2018: 4038–4041

53 Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015;7(299):299ra122

54 Brown SM, Jones J, Kuttler KG, Keddington RK, Allen TL, Haug P. Prospective evaluation of an automated method to identify patients with severe sepsis or septic shock in the emergency department. BMC Emerg Med 2016;16(01):31

55 Gunnarsdottir K, Sadashivaiah V, Kerr M, Santaniello S, Sarma SV. Using demographic and time series physiological features to classify sepsis in the intensive care unit. Conf Proc IEEE Eng Med Biol Soc 2016;2016:778–782

56 Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. Comput Biol Med 2016; 74:69–73

57 Desautels T, Calvert J, Hoffman J, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 2016;4 (03):e28

58 Mao Q, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018;8(01):e017833

59 Rech MA, Mosier MJ, Zelisko S, Netzer G, Kovacs EJ, Afshar M. Comparison of automated methods versus the American Burn Association sepsis definition to identify sepsis and sepsis with organ dysfunction/septic shock in burn-injured adults. J Burn Care Res 2017;38(05):312–318

60 Sorsa A, Früh J, Stötter L, Abdissa S. Blood culture result profile and antimicrobial resistance pattern: a report from neonatal intensive care unit (NICU), Asella teaching and referral hospital, Asella, South East Ethiopia. Antimicrob Resist Infect Control 2019;8:42

61 Creedon JK, Vargas S, Asaro LA, Wypij D, Paul R, Melendez E. Timing of antibiotic administration in pediatric sepsis. Pediatr Emerg Care 2018. Nov 26, 2018. doi: 10.1097/PEC.0000000000001663. [Epub ahead of print]

62 Dellinger RP, Levy MM, Carlet JM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Intensive Care Med 2008;34(01):17–60

63 Opota O, Jaton K, Greub G. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. Clin Microbiol Infect 2015;21(04):323–331

64 Miller JM, Binnicker MJ, Campbell S, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2018 update by the Infectious Diseases Society of America and the American Society for Microbiology. Clin Infect Dis 2018;67(06): e1–e94

65 Schuh CJ. Sepsis and septic shock analysis using neural networks. Paper presented at: Annual Meeting of the North American Fuzzy Information Processing Society; June 24, 2007; San Diego, CA, United States

66 Harrison AM, Thongprayoon C, Kashyap R, et al. Developing the surveillance algorithm for detection of failure to recognize and treat severe sepsis. Mayo Clin Proc 2015;90(02):166–175

67 Nelson JL, Smith BL, Jared JD, Younger JG. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. Ann Emerg Med 2011;57(05):500–504

68 Nachimuthu SK, Haug PJ. Early detection of sepsis in the emergency department using Dynamic Bayesian Networks. AMIA Annu Symp Proc 2012;2012:653–662

69 Brink A, Alsma J, Verdonschot RJCG, et al. Predicting mortality in patients with suspected sepsis at the emergency department; a retrospective cohort study comparing qSOFA, SIRS and National Early Warning Score. PLoS One 2019;14(01):e0211133

70 Wu Z, Liang Y, Li Z, et al. Accuracy comparison between age-adapted sofa and sirs in predicting in-hospital mortality of infected children at China's PICU. Shock. Sep 11, 2018. doi: 10.1097/SHK.0000000000001261. [Epub ahead of print]

71 Gaini S, Relster MM, Pedersen C, Johansen IS. Prediction of 28-days mortality with sequential organ failure assessment (SOFA), quick SOFA (qSOFA) and systemic inflammatory response syndrome (SIRS) - a retrospective study of medical patients with acute infectious disease. Int J Infect Dis 2019;78:1–7

72 Thompson DS, Oberteuffer R, Dorman T. Sepsis alert and diagnostic system: integrating clinical systems to enhance study coordinator efficiency. Comput Inform Nurs 2003;21(01):22–26, quiz 27–28

73 Giuliano KK. Physiological monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. Am J Crit Care 2007;16(02):122–130, quiz 131

74 Parente JD, Lee D, Lin J, Chase JG, Shaw GM. A fast and accurate diagnostic test for severe sepsis using kernel classifiers. Paper presented at: UKACC International Conference on Control; September 7, 2010; Coventry, United Kingdom

75 Thiel SW, Rosini JM, Shannon W, Doherty JA, Micek ST, Kollef MH. Early prediction of septic shock in hospitalized patients. J Hosp Med 2010;5(01):19–25

76 Amland RC, Lyons JJ, Greene TL, Haley JM. A two-stage clinical decision support system for early recognition and stratification of patients with sepsis: an observational cohort study. JRSM Open 2015;6(10):2054270415609004

77 Brandt BN, Gartner AB, Moncure M, et al. Identifying severe sepsis via electronic surveillance. Am J Med Qual 2015;30(06): 559–565

78 Amland RC, Haley JM, Lyons JJ. A multidisciplinary sepsis program enabled by a two-stage clinical decision support system: factors that influence patient outcomes. Am J Med Qual 2016;31(06):501–508

79 Amland RC, Hahn-Cover KE. Clinical decision support for early recognition of sepsis. Am J Med Qual 2016;31(02):103–110

80 Back J-S, Jin Y, Jin T, Lee S-M. Development and validation of an automated sepsis risk assessment system. Res Nurs Health 2016; 39(05):317–327

81 Rolnick J, Downing NL, Shepard J, et al. Validation of test performance and clinical time zero for an electronic health record embedded severe sepsis alert. Appl Clin Inform 2016;7(02):560–572

82 Zuick S, Asch DA, Graustein A, Urbani R, Carrol B. Can a computerized sepsis screening and alert system accurately diagnose sepsis in hospitalized floor patients and potentially provide

opportunities for early intervention? A pilot study. J Intensiv Crit Care 2016;02:03

83 Gille J, Dietz A, Taha H, Sablotzki A. A sirs-based automated alarm system for the diagnosis of sepsis after burn injury. Ann Burns Fire Disasters 2017;30(03):177–184

84 Martin Rico P, Valdivia Pérez A, Lacalle Martínez JM, et al. Electronic alerting and decision support for early sepsis detection and management: Impact on clinical outcomes. Eur J Clin Pharmacol 2017;19(01):33–40

85 Olenick EM, Zimbro KS, D'Lima GM, Ver Schneider P, Jones D. Predicting sepsis risk using the "Sniffer" algorithm in the electronic medical record. J Nurs Care Qual 2017;32(01):25–31

86 Shashikumar SP, Stanley MD, Sadiq I, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. J Electrocardiol 2017;50(06):739–743

87 Westra BL, Landman S, Yadav P, Steinbach M. Secondary analysis of an electronic surveillance system combined with multi-focal interventions for early detection of sepsis. Appl Clin Inform 2017; 8(01):47–66

88 Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. J Am Med Inform Assoc 2018;25 (05):523–529

89 Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. Ann Emerg Med 2019; 73(04):334–344

90 Faisal M, Scally A, Richardson D, et al. Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. Crit Care Med 2018;46(04):612–618

91 Khoshnevisan F, Ivy J, Capan M, Arnold R, Huddleston J, Chi M. Recent temporal pattern mining for septic shock early prediction. Paper presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 4, 2018; New York, NY, United States

92 Thakur J, Pahuja SK, Pahuja R. Neonatal sepsis prediction model for resource-poor developing countries. Paper presented at: 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech); May 4, 2018; Kolkata, India

93 van Wyk F, Khojandi A, Kamaleswaran R. Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study. IEEE J Biomed Health Inform 2019;23(03): 978–986

94 van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. Int J Med Inform 2019;122:55–62