Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review [version 2; peer review: 2 approved]

Goran Medic1,2, Melodi Kosaner Kließ3, Louis Atallah4, Jochen Weichert4, Saswat Panda3, Maarten Postma2,5,6, Amer EL-Kerdi4

1Health Economics, Philips, Eindhoven, Noord-Brabant, 5621JG, The Netherlands
2Department of Pharmacy, Unit of PharmacoTherapy, -Epidemiology & -Economics, University of Groningen, Groningen, 9700 AB, The Netherlands
3Global Market Access Solutions Sàrl, St-Prex, 1162, Switzerland
4Philips, Cambridge, MA, 02141, USA
5Department of Health Sciences, University Medical Centre Groningen, University of Groningen, Groningen, 9700 AB, The Netherlands
6Department of Economics, Econometrics & Finance, University of Groningen, Groningen, 9700 AB, The Netherlands

Abstract

Background: Clinical decision support (CDS) systems have emerged as tools providing intelligent decision making to address challenges of critical care. CDS systems can be based on existing guidelines or best practices; and can also utilize machine learning to provide a diagnosis, recommendation, or therapy course.

Methods: This research aimed to identify evidence-based study designs and outcome measures to determine the clinical effectiveness of clinical decision support systems in the detection and prediction of hemodynamic instability, respiratory distress, and infection within critical care settings. PubMed, ClinicalTrials.gov and Cochrane Database of Systematic Reviews were systematically searched to identify primary research published in English between 2013 and 2018. Studies conducted in the USA, Canada, UK, Germany and France with more than 10 participants per arm were included.

Results: In studies on hemodynamic instability, the prediction and management of septic shock were the most researched topics followed by the early prediction of heart failure. For respiratory distress, the most popular topics were pneumonia detection and prediction followed by pulmonary embolisms. Given the importance of imaging and clinical notes, this area combined Machine Learning with image analysis and natural language processing. In studies on infection, the most researched areas were the detection, prediction,
and management of sepsis, surgical site infections, as well as acute kidney injury. Overall, a variety of Machine Learning algorithms were utilized frequently, particularly support vector machines, boosting techniques, random forest classifiers and neural networks. Sensitivity, specificity, and ROC AUC were the most frequently reported performance measures.

**Conclusion:** This review showed an increasing use of Machine Learning for CDS in all three areas. Large datasets are required for training these algorithms; making it imperative to appropriately address challenges such as class imbalance, correct labelling of data and missing data. Recommendations are formulated for the development and successful adoption of CDS systems.

**Keywords**
sepsis, hemodynamic instability, respiratory distress, infection, machine learning, clinical trials, critical care.

---

**Corresponding author:** Goran Medic (goran.medic@philips.com)

**Author roles:**
- **Medic G:** Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation
- **Kosaner Kließ M:** Data Curation, Formal Analysis, Methodology, Project Administration, Validation, Writing – Review & Editing
- **Atallah L:** Writing – Original Draft Preparation, Writing – Review & Editing
- **Weichert J:** Writing – Review & Editing
- **Panda S:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing
- **Postma M:** Conceptualization, Supervision, Writing – Review & Editing
- **EL-Kerdi A:** Conceptualization, Funding Acquisition, Methodology, Supervision, Validation, Writing – Review & Editing

**Competing interests:** PM has no conflicts of interest. MG, AL, WJ and ELKA are the employees of Philips. KKM and PS are the employees of Global Market Access Solutions Sàrl. Received funding from Philips to perform systematic literature review. PM is the employee of the University of Groningen, The Netherlands who provided scientific oversight for the whole project and did not receive any financial support.

**Grant information:** The study was supported by funding from Philips. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Copyright:** © 2019 Medic G et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Medic G, Kosaner Kließ M, Atallah L et al. Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review [version 2; peer review: 2 approved] F1000Research 2019, 8:1728 https://doi.org/10.12688/f1000research.20498.2

**First published:** 08 Oct 2019, 8:1728 https://doi.org/10.12688/f1000research.20498.1
Amendments from Version 1

All comments from the Reviewers were addressed in the updated version. We could not address the layout issue that Reviewer 1 made as this is the Journal’s decision how tables are made in the PDF.

The question of Reviewer 2 regarding the rationale for including the studies predicting AKI within the Infection/sepsis results section is addressed here:

Severe infection is a major cause of AKI in ICU patients, while conversely, AKI patients are at increased risk for infection [1]. Sepsis is an important cause of AKI, and AKI is a common complication of sepsis [2]. We felt that given this relationship, CDS for AKI fits well under this section. The reviewer is correct to propose the link between AKI and shock; however, not all AKI cases lead to shock—so we felt it matched this section more.

[1] Vandijck DM, Reynvoet E, Blot SI, Vandecasteele E, Hoste EA. Severe infection, sepsis and acute kidney injury. Acta Clin Belg. 2007;62 Suppl 2:332-6.

[2] Steven J. Skube, Stephen A. Katz, Jeffrey G. Chipman, and Christopher J. Tignanelli. Surgical Infections. http://doi.org/10.1089/sur.2017.26 Sur.2017.26

Additional references

[3] Vandijck DM, Reynvoet E, Blot SI, Vandecasteele E, Hoste EA. Severe infection, sepsis and acute kidney injury. Acta Clin Belg. 2007;62 Suppl 2:332-6.

[4] Steven J. Skube, Stephen A. Katz, Jeffrey G. Chipman, and Christopher J. Tignanelli. Surgical Infections. http://doi.org/10.1089/sur.2017.26 Sur.2017.26

Introduction

Critical care, including intensive and emergency care, is the most expensive and human resource intensive area of in-hospital care. Despite having the most technologically advanced devices, it is the area associated with the highest morbidity and mortality rates. Decision-making for clinical teams in this area is complex due to variability in procedures and data-overload from the plethora of existing devices. In fact, misdiagnosis in the intensive care unit (ICU) is 50% more common than other areas1, and errors, especially medication errors which account for 78% of serious medication errors, can have a long lasting effect even after patients are discharged.

Computerized decision support (CDS) systems have emerged as tools providing intelligent decision making based on patient data to address many of the challenges of critical care. CDS systems can be based on existing guidelines or best practices; and can also utilize machine learning as a means of compiling several data inputs to provide a diagnosis, recommendation, or therapy course. CDS systems can improve medication safety by providing recommendations relating to dosing2, administration frequencies3, medication discontinuation4, and medication avoidance5. Moreover, these novel systems can improve the quality of prescribing decisions by triggering alerts or warning messages on drug duplication, contraindications, drug interaction errors, side-effects and inappropriate medication orders. CDS system notifications can be applied during the prescribing, administering or monitoring stages to detect and prevent medication errors. These systems can also target patients to facilitate shared decision-making to empower as well as to motivate them. The need for such systems stems from hospitals having to deal with strict guidelines to improve outcomes, document care cycles (raising the need for administrative tasks) and reduce readmissions. This is combined with the need to cope with financial constraints, such as staff shortages and increased pressure to reduce the length of stay5,6.

Strategies for bringing CDS to clinics have been the topic of several workshops, conferences and focus groups7. Factors for success in designing CDS include providing measurable value, producing actionable insights, delivering information to the user at the right time, and demonstrating good usability principles7.

Early warning systems (EWS) are CDS systems designed for initial assessment and identification of patients at risk of deterioration in in-patient ward areas8,9. These systems have shown that they can enable caregivers and rapid response teams to respond earlier—in time to make a difference10. By alerting clinicians to higher risk patients, treatments can be administered early or harmful medications can be stopped, potentially leading to improved outcomes. Early recognition and timely intervention are also critical steps for the successful management of shock11, cardiopulmonary instability12 and severe sepsis. In sepsis management, adequate timing of administration of antibiotics is directly associated with survival rates13, and incidence, severity and duration of infections.

According to the Society of Critical Care Medicine (SCCM)14, the five primary ICU admission diagnoses for adults are respiratory insufficiency/failure with ventilator support, acute myocardial infarction, intracranial hemorrhage or cerebral infarction, percutaneous cardiovascular procedures, and septicemia or severe sepsis without mechanical ventilation. SCCM also highlights other conditions including high ICU demand such as poisoning and toxic effects of drugs, pulmonary edema and respiratory failure, heart failure and shock, cardiac arrhythmia and renal failure. Given the above, three high-impact areas were selected for the current research where early detection and treatment could impact outcomes for patients in the ICU. The first is that of hemodynamic instability, where early detection could help patients prevent deterioration into shock. The second is that of respiratory distress, affecting many ventilated patients (up to 40% are ventilated according to SCCM15). The third area selected is that of infection, with a focus on sepsis. Sepsis is the most common cause of death among critically ill patients, with occurrence rates varying from 13.6% to 39.3%16,17. All three areas are major areas of concern with relatively high prevalence in critical care having long term effects on patients.

The study focuses on both detection, which alerts the clinician to the presence of these specific conditions, as well as prediction of deterioration by alerting the clinician in advance that a patient will deteriorate into one of these disease states. The aims of this study were to perform and report a systematic review of the utilization of CDS systems in the three selected disease areas and summarize the methodological aspects of identified studies.

Methods

Search strategy

A systematic literature review was carried out to identify evidence-based study designs, methods and outcome measures that have been used to determine the clinical effectiveness of CDS systems in the detection and prediction of three populations representing the variety and majority of morbid conditions in a critical care setting: Shock (hemodynamic (in-)stability),
Table 1. Study selection criteria for the systematic literature review.

| Criteria          | Inclusion                                                                                     | Exclusion                                                                                           |
|-------------------|----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| **STUDY DESIGN**  | **Abstract selection**                                                                        | Systematic Literature Reviews or meta-analyses*                                                     |
|                   | Randomized controlled trials (RCT)                                                            | Review papers, newsletters and opinion papers where treatments of interest are only discussed      |
|                   | Observational (retrospective and prospective) studies                                          | Methodology studies or protocols                                                                    |
|                   | In-hospital settings: Acute care, Intensive care unit (ICU), Emergency department (ED), Medical Surgery, General ward | Case studies (sample size of 1 patient)                                                              |
|                   | Geography: US, Canada, Europe                                                                 | Studies with less than 10 patients per arm;                                                        |
|                   |                                                                                              | Conference abstracts published only as abstracts in 2013, 2014, 2015 and 2016                     |
|                   |                                                                                              | Geography**: All countries and regions except: US, Canada, UK, Germany, France                       |
|                   | **Full-text selection**                                                                       | Publications without an abstract                                                                    |
|                   | Randomized controlled trials (RCT)                                                            | Systematic Literature Reviews or meta-analyses*                                                     |
|                   | Observational (retrospective and prospective) studies                                          | Review papers, newsletters and opinion papers where treatments of interest are only discussed      |
|                   | In-hospital settings: Acute care, Intensive care unit (ICU), Emergency department (ED), Medical Surgery, General ward | Methodology studies or protocols                                                                    |
|                   | Geography**: US, Canada, UK, Germany, France                                                 | Case studies (sample size of 1 patient)                                                              |
|                   | Conference abstracts published only as abstracts in 2017 and 2018                            | Studies with less than 10 patients per arm;                                                        |
|                   |                                                                                              | Geography**: All countries and regions except: US, Canada, UK, Germany, France                       |
| **POPULATION**    | **Abstract and full-text selection**                                                          | **In-vitro studies**                                                                                 |
|                   | Studies that include humans only – adults, children and neonates (or electronic) medical records | Animal studies                                                                                      |
|                   | Both sexes are included Patients with or at risk of developing shock (hemodynamic instability) |                                                                                                     |
|                   | Patients with or at risk of developing respiratory distress/failure                           |                                                                                                     |
|                   | Patients with or at risk of developing infection or sepsis                                   |                                                                                                     |
|                   | Healthy people only; Healthy people and patients                                             |                                                                                                     |
### Criteria

| Criteria / Intervention | Inclusion | Exclusion |
|-------------------------|-----------|-----------|
| **TREATMENT / INTERVENTION** | Abstract and full-text selection | Artificial intelligence<br>Machine learning (i.e. Deep learning models)<br>Clinical decision support<br>Computer aided detection<br>Early Warning System | Automatic diagnosis systems (i.e. ELISA tests)<br>Screening tests (i.e. Automated analysis of portable oximetry)<br>Sequencing tests<br>Mathematical models*** - which model the predictability of disease or treatment/ intervention (i.e. Modelling studies have been widely used to inform human papillomavirus vaccination policy decisions)<br>Multivariable hierarchal logistic regression models*** (models which are based only on statistics - but there is no machine learning) |
| **COMPARATOR** | Abstract and full-text selection | All comparators | No selection will be made regarding comparator |
| **OUTCOMES** | Abstract and full-text selection | Detection and/or prediction outcomes, such as:<br>• Sensitivity (SD) (%)<br>• Specificity (SD) (%)<br>• NPV (%)<br>• PPV (%)<br>• Likelihood ratio<br>• Accuracy (SD) (%)<br>• Prevalence of disease (%)<br>• OR; 95% CI; p-value<br>• HR; 95% CI; p-value<br>• Median (IQR); p-value<br>• ROC AUC | Studies not reporting detection and/or prediction outcomes<br>Studies discussing interventions of interest, but no outcomes are reported |

* Systematic Literature Reviews and (network) meta-analysis are excluded from data extraction since the pooled results cannot be used in our analysis. However, good quality (network) meta-analysis and systematic literature reviews (i.e. Cochrane reviews) will be used for cross-checking of references if the search did not omit any articles.

** If studies are conducted in multiple countries and at least 1 of the included countries is included – the study will be included in the selection.

*** Mathematical and logistic regression models – can be used to validate and evaluate Interventions of interest (that are listed as included intervention), but the texts discussing these models without any “learning potential” or artificial intelligence potential will be excluded. Therefore, these models can be the foundation of the included listed interventions but will not be included in the Data Extraction Files unless they have also machine learning or artificial intelligence or some other form of “learning potential” on top of the statistical mathematical model. Researchers will pay special attention and caution when screening these abstracts and/or full-text articles.

AUC = Area under the curve; ED = Emergency department; ELISA = Enzyme-linked immunosorbent assay; HR = Hazard ratio; ICU = Intensive care unit; IQR = interquartile range; NPV = Negative predictive value; OR = Odds ratio; PPV = Positive predictive value; RCT = Randomized controlled trial; ROC = Receiver Operating Characteristic; SD = Standard deviation; SE = Standard error; UK = United Kingdom; US = United States.

### Studies identified from the ClinicalTrials.gov registry that did not report results were also included in the extraction to give some indication of the outcomes being collected.

### Study quality appraisal

This research was not aimed at summarizing study results and assessing the relative effectiveness of CDS systems. Therefore, an appraisal of study quality was not deemed necessary.
Results

Shock (hemodynamic (in-)stability)
The search yielded 1588 hits. Screening the titles and abstracts led to 1502 being excluded. The full texts of the remaining 86 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=22), population (n=1), intervention (n=5), and outcomes (n=38). A total of 20 studies were finally included in this systematic literature review. This included 5 trials identified from ClinicalTrials.gov. The study selection process is depicted in Figure 1.

Study characteristics. Of the 15 published studies, five were conducted by research groups outside the USA. Ten studies were conducted in the US. Thirteen studies were retrospective, and only two were prospective. Nine studies were single-center and six studies were multi-center. Five studies were time-series and nine were case-series.

Across all studies, three had sample sizes ≤100; three had sample sizes of 101–1000; four had sample sizes of 1001–10,000; and another five studies, four retrospective single-center studies and one multi-center, had sample sizes larger than 10,000. The three largest studies included patients admitted to various wards of a specified hospital. The majority of the studies did not restrict their sample to a specific in-patient hospital setting. Five studies reported on patients...
in the ICU\textsuperscript{39,28,32,40,41} and one study reported on patients admitted to the surgical ward\textsuperscript{42}.

The characteristics of the published studies are summarized in Table 2.

\textbf{CDS systems.} Machine learning algorithms were developed to detect or predict septic shock\textsuperscript{28,33,35,40,41}, various heart arrhythmias\textsuperscript{29,30,34}—heart failure\textsuperscript{37–39}—hemodynamic instability and hypovolemia\textsuperscript{30,36}, myocardial infarction\textsuperscript{31}, as well as hypotension\textsuperscript{15}.

All studies, except one, trained a single algorithm. Ebrahimzadeh et al.\textsuperscript{18} trained and compared support vector machine (SVM), instance-based and neural network models to predict paroxysmal atrial fibrillation. SVMs were the most frequently used algorithms, followed by least absolute shrinkage and selection operator (LASSO) regularization. In one study, the SVM was trained using sequential minimal optimization\textsuperscript{16}.

Machine learning models were trained and validated in 14 studies and subsequently tested in an independent dataset in 3 studies\textsuperscript{35,30,37}. In one study an algorithm trained to classify arrhythmias was not validated but compared to physician’s manual classifications\textsuperscript{34}.

An overview of the investigated machine learning algorithms is presented in Table 3.

\textbf{Outcome measures.} Three of the 15 papers measured a single outcome of model performance. In two studies the preferred measure was accuracy\textsuperscript{28,34}; whereas in another study this was the ROC AUC. This study was large and based their algorithm on EHRs\textsuperscript{31}. Across all studies, accuracy was reported in about half of the instances and the ROC AUC was one of the most frequently reported outcomes.

Sensitivity and specificity were reported together in 10 studies. Blecker et al.\textsuperscript{18} reported sensitivity together with PPV. Sensitivity and specificity were not measured in the study by Sideris et al.\textsuperscript{17}, instead model accuracy and the ROC AUC were preferred. This study was concerned with developing an alternative ‘comorbidity’ framework based on disease and symptom diagnostic codes to cluster individuals at low to high risk of developing chronic heart failure.

PPVs were reported in six studies and accompanied with negative predictive values in two studies. These studies developed and validated machine-learning algorithms for the early detection of less investigated health conditions, these being hemodynamic instability in children\textsuperscript{39} and acute decompensated heart failure\textsuperscript{39}. The highest number of outcome measures, including likelihood ratios, was observed in Calvert et al.\textsuperscript{16} who investigated an under-represented population of patients with Alcohol Use Disorder.

The outcomes measured are summarized in Table 4.

\textbf{Ongoing studies.} Five studies are currently ongoing, one in Germany\textsuperscript{43} and the others in the USA\textsuperscript{44–47}. Two studies are prospective case series\textsuperscript{44,47}, two studies are prospective cohort studies\textsuperscript{44,47} and one is a RCT\textsuperscript{44}. Two of the studies are concerned with developing prediction models, and the others are concerned with implementing machine learning algorithms into clinical practice as early warning systems.

The details of these trials are summarized in Table 5.

\textbf{Respiratory distress/failure}

The search yielded 1279 hits. Screening the titles and abstracts lead to 1142 being excluded. The full texts of the remaining 137 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=42), population (n=6); intervention (n=18) and outcomes (n=47), and conference proceeding from before 2017 (n=2). A total of 22 studies were finally included in this systematic literature review. None of the trials retrieved from ClinicalTrials.gov were included. The study selection process is depicted in Figure 2.

\textbf{Study characteristics.} Of the included studies, 17 were conducted in the US\textsuperscript{31,46–63}. Five studies were conducted outside the US; two in Canada\textsuperscript{34,65} by the same research group, two in France\textsuperscript{66,67} and one in the UK\textsuperscript{68}. In total, 17 studies were retrospective\textsuperscript{33,48–50,52–55,58–66} and five were prospective\textsuperscript{51,56,57,67,68}. Of these studies, 12 were single-center\textsuperscript{33,48,49,51,52,54,55,58,59,63–66} and 10 studies were multi-center\textsuperscript{50,53,56,60–63,67,68}. Five studies were time-series\textsuperscript{46,52,55,56,64}, 14 studies were case-series\textsuperscript{33,40,41,53,54,57–59,62–65,66,68}, one was case-control\textsuperscript{10} and one was case/time series study\textsuperscript{31}.

The smallest sample of 100 patients came from two single-center retrospective studies\textsuperscript{46,66}. Ten studies had sample sizes of 101–1000\textsuperscript{33,39,34–37,57,83,84}, seven studies had sample sizes of 1001–10,000\textsuperscript{35,59,60,62,64,83}, and three had sample sizes larger than 10,000\textsuperscript{65,86,91}. The largest study included more than 50,000 patients admitted to the ED of two centers over a 3-year period\textsuperscript{64}. Several published studies did not report their in-patient setting. When reported, some evaluated data from different wards\textsuperscript{46,52}, and some included patients admitted only to the ED\textsuperscript{33,34,61,65}, the ICU\textsuperscript{50,60,67} and the surgical ward\textsuperscript{51}.

The characteristics of all published studies are given in Table 6.

\textbf{CDS systems.} About half of the studies developed machine-learning algorithms, whereas the other half focused on natural language processing (NLP) algorithms. One study differed from the rest by developing a computer-aided detection (CAD) system to measure the axial diameter of the right and left pulmonary ventricles, aiding in the diagnosis of pulmonary embolisms\textsuperscript{49}. Many learning algorithms were concerned with detecting pulmonary embolisms and deep vein thrombosis\textsuperscript{33,34,58,59,64–67} as well as pneumonia\textsuperscript{53,48,57,60–63}. Three studies developed machine-learning algorithms to detect COPD\textsuperscript{60,64,69}. One study developed a machine learning algorithm to detect acute respiratory distress syndrome\textsuperscript{52}; while other studies developed machine learning algorithms to detect respiratory distress or failure following a pressure support ventilation trial\textsuperscript{67}, cardiovascular surgery\textsuperscript{91} and pediatric tonsillectomy\textsuperscript{91}.
| Study       | Study Design            | Country and institution(s)                                                                 | Number of patients (records) | Population/disease definition | In-patient setting | Collected data                                                                 |
|------------|-------------------------|-------------------------------------------------------------------------------------------|------------------------------|------------------------------|-------------------|--------------------------------------------------------------------------------|
| Ghosh 2017 | Retrospective time series single center | Australia University of Technology Sydney & The University of Melbourne                     | 209                          | Sepsis or severe sepsis      | ICU               | (mean arterial pressure), heart rate, respiratory rate                          |
| Hu 2016    | Retrospective case series single center | USA, Minnesota University of Minnesota                                                      | NR (8909)                    | NR                           | Surgery           | EHRs                                                                            |
| Li 2014    | Retrospective case series multi-centric (3 centers) | UK, Oxford University of Oxford & Mindray                                                   | NR (67)                      | Ventricular flutter, fibrillation and tachycardia | NR                | Electrocardiography                                                              |
| Mahajan 2014 | Prospective case series multi-centric (4 centers) | USA University of Southern California, Mayo Clinic-Rochester, University of North Carolina, Sanger Heart & Vascular Institute & Boston Scientific | 410 (908)                    | Ventricular fibrillation, ventricular tachycardia and other arrhythmias | NR                | Electrograms                                                                    |
| Mao 2018   | Retrospective case series multi-centric (5 centers) | USA University of California, Stanford Medical Centre, Oroville Hospital, Bakersfield Heart Hospital, Cape Regional Medical Centre, Beth Israel Deaconess Medical Center | 359,390                      | NR                           | various           | Vital signs                                                                     |
| Reljin 2018 | Prospective case-control multi-centric (2 centers) | USA University of Connecticut, Campbell University School of Medicine, University of Massachusetts Medical School, Yale University School of Medicine & Worcester Polytechnic Institute | 36 (94)                      | Traumatic injury, healthy controls | NR                | Photoplethysmographic signals                                                    |
| Sideris 2016 | Retrospective case series single center | USA, Los Angeles University of California                                                   | 1948                         | Primarily heart failure      | various           | EHRs                                                                            |
| Blecker 2016 | Retrospective case series single center | USA, New York NewYork-Presbyterian Hospital & New York University                          | NR (47,119)                  | NR                           | various           | EHRs                                                                            |
| Blecker 2018 | Retrospective case series single center | USA, New York New York University                                                             | NR (37229)                   | NR                           | various           | EHRs                                                                            |
| Study          | Study Design                              | Country and institution(s)                                                                 | Number of patients (records) | Population/disease definition | In-patient setting | Collected data                                               |
|---------------|-------------------------------------------|-------------------------------------------------------------------------------------------|------------------------------|-----------------------------|-------------------|--------------------------------------------------------------|
| Calvert 2016  | Retrospective time series single center   | USA, California Dascena Inc. & University of California                                   | 29083                       | NR                          | ICU               | Vital signs                                                  |
| Donald 2018   | Retrospective time series + Prospective time series multi-centric (22 centers) | Europe                                                                                  | 173                         | Traumatic brain injury      | ICU               | Demographic, clinical and physiological data                |
| Ebrahimzadeh 2018 | Retrospective time series single center | Iran University of Tehran, Iran University of Science and Technology, University of Sheikhbahaee & Payame Noor University of North Tehran | 53 (106)                     | Paroxysmal atrial fibrillation | NR                | Electrocardiography                                          |
| Potes 2017    | Retrospective case series multi-centric (2 centers) | USA, California & UK, London Children’s Hospital Los Angeles, St. Mary’s Hospital, London & Philips | 8022                        | NR                          | ICU               | Vital signs, laboratory values, and ventilator parameters.   |
| Henry 2015    | Retrospective case series single center   | USA; Maryland John Hopkins University                                                      | 16234                       | NR                          | ICU               | EHRs                                                        |
| Strodthoff 2018 | Retrospective time series single center | Germany, Berlin Fraunhofer Heinrich Hertz Institute & University Medical Center Schleswig-Holstein, Kiel | 200 (228)                     | Myocardial infarction and healthy controls | NR                | Electrocardiography                                          |

USA: United States of America. UK: United Kingdom. NR: Not reported. ICU: Intensive care unit. EHR: Electronic health records.

The classifiers used in the NLP-based studies were various. However, some commonalities emerged between the studies developing machine-learning algorithms. Multiple studies applied SVM, logistic regression, random forests, K-nearest neighbor (kNN), gradient boosting and neural network models. Various classifiers were explored in 5 studies.

Machine learning and NLP-based algorithms were trained and validated in 20 studies and subsequently tested in an independent dataset in 6 studies. The CAD system mentioned above and an electronic pulmonary embolism severity index were trained and compared to a reference dataset classified by physicians.

An overview of the developed learning algorithms is provided in Table 7.

One study, Reamoroon et al. 2018, used a novel sampling technique to accommodate for inter-dependency in longitudinal data. Model accuracy and ROC AUC with this method was <5% better than random sampling and 4–11% better than no sampling.

**Outcome measures.** The majority of the studies reported multiple outcome measures of model performance. The most frequently reported outcome measure was sensitivity, followed by specificity and ROC AUC. Likelihood ratios, on the other hand, were only reported in one study: Silva et al. 2017 reported eight outcome measures of their novel machine learning model to predict post extubation distress. The outcomes measured across all studies are summarized in Table 8.

Many of the studies that developed NLP-based algorithms reported negative and positive predictive values, as well as sensitivity and specificity. In contrast, the ROC AUC was the most frequently reported outcome measure of machine learning.
### Table 3. Overview of the algorithms developed to detect shock.

| Study        | Predicted disease                | CHMM | Decision trees | LR LASSO | LR, not specified | SVM | kNN | RF | gradient tree boosting | Adaptive boosting | Bayesian neural network | convolutional neural network | Multilayer perception | mixture of expert |
|--------------|----------------------------------|------|----------------|----------|------------------|-----|-----|----|------------------------|------------------|------------------------|------------------------|----------------------|---------------------|
| Ebrahimzadeh 2018 | paroxysmal atrial fibrillation | ✓    | ✓              |          | ✓                |     |     |    |                        |                  |                        |                        |                      |                     |
| Li 2014      | Ventricular fibrillation and tachycardia | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Mahajan 2014 | heart arrhythmias                | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Strothoff 2018 | myocardial infarction          |      |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Sideris 2016 | heart failure                    | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Blecker 2016 | heart failure                    | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Blecker 2018 | heart failure                    | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Reljin 2018  | Hypovolemia                      |      |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Potes 2017   | hemodynamic instability          | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Donald 2018  | Hypotension                      | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Ghosh 2017   | septic shock                     |      |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Hu 2016      | septic shock                     | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Mao 2018     | septic shock                     | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Calvert 2016 | septic shock                     |      |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |
| Henry 2015   | septic shock                     | ✓    |                |          |                  |     |     |    |                        |                  |                        |                        |                      |                     |

CHMM: clustered hidden Markov model. LR: Logistic regression. SVM: Support vector machine. kNN: k nearest neighbor. RF: Random forest. Conv.: Convolutional.

### Table 4. Overview of measured outcomes in studies on shock.

| Study        | Sensitivity | Specificity | NPV | PPV | Negative LR | Positive LR | Accuracy | Prevalence | OR | RR | ROC AUC |
|--------------|-------------|-------------|-----|-----|-------------|-------------|----------|------------|----|----|---------|
| Ghosh 2017   |             |             | ✓   | ✓   |             |             |          |            |    |    |         |
| Hu 2016      |             |             | ✓   | ✓   |             |             |          |            |    |    |         |
| Li 2014      | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Mahajan 2014 |             |             | ✓   | ✓   |             |             |          |            |    |    |         |
| Mao 2018     | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Reljin 2018  | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Sideris 2016 |             |             | ✓   | ✓   |             |             |          |            |    |    |         |
| Blecker 2016 | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Blecker 2018 | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Calvert 2016 | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Donald 2018  | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Ebrahimzadeh 2018 | ✓     | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Potes 2017   | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Henry 2015   | ✓           | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |
| Strothoff 2018 | ✓     | ✓           | ✓   | ✓   |             |             |          |            |    |    |         |

NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio. RR: Risk ratio. ROC AUC: Receiver operating characteristic area under the curve.
## Table 5. Overview of ongoing studies on shock.

| Identifier code | Study Design | Countries and study centers | Hospital setting | Intervention | Sample characteristics | Outcome(s) |
|-----------------|--------------|-----------------------------|------------------|--------------|------------------------|------------|
| NCT03582501    | Prospective case series | USA, Mayo Clinic, Arizona, Florida & Rochester | NR | Lower body negative pressure to simulate hypovolemia | Estimated: 24 Age: 18–55 Definition: Healthy non-smoker, no history of hypertension, diabetes, CAD and neurologic diseases | Primary outcome Blood pressure Secondary outcome Heart rate |
| NCT02934971    | Prospective cohort study | Germany, Aachen University Hospital | Out-patient | Chemotherapy or no chemotherapy | Estimated: 400 Age: ≥ 18 Definition: Patients scheduled for chemotherapy at increased risk of cardiotoxicity and age-matched controls | Primary outcome change in left ventricular ejection fraction |
| NCT03235193    | Prospective cohort study | USA, West Virginia Dascena Inc. & University of California | ED, ICU | The InSight algorithm used as an EWS to detect sepsis and severe sepsis detection from EHRs compared to severe sepsis detection from EHRs alone | Estimated: 1241 Age: ≥ 18 Definition: All admitted patients | Primary outcome in-hospital mortality Secondary outcomes length of stay in hospital and ICU, hospital readmission |
| NCT03644940    | RCT | USA, California Dascena Inc. & University of California | Cardiology, GI, ICU, Medicine, Oncology, Surgery, Transplant and ED | subpopulation-optimized version of InSight compared to the original version used as an early warning system to identify patients at high risk of severe sepsis; followed by physician assessment of sepsis | Estimated n: 51645 Age: >18 Definition: NR | Primary outcomes in-hospital SIRS-based mortality Secondary outcomes in-hospital severe sepsis/shock-coded mortality; SIRS-based hospital length of stay; Severe sepsis/shock-coded hospital length of stay |
| NCT03655626    | Single-arm trial up to | USA, North Carolina Duke University Hospital | ED | machine learning algorithm to predict sepsis, custom dashboard and monitoring | Estimated n: 3200 Age: >18 Definition: NR | Primary outcome rate of CMS bundle completion for patients with sepsis Secondary outcomes time to sepsis diagnosis; number of patients developing sepsis; number of patients developing sepsis and not treated; length of stay in ED and hospital; inpatient mortality; ICU requirement rate; time from sepsis onset to blood culture, antibiotics, IV fluids, lactate, CMS bundle completion; rate of lactate complete; number of sepsis diagnostic codes per month |

USA: United States of America. NR: Not reported. ED: Emergency department. ICU: Intensive care unit. GI: Gastroenterology.
Figure 2. Study selection - Respiratory distress-failure. Pop. = Population.
| Study        | Study Design                      | Countries and institution(s)                                      | Number of patients (records) | Population/disease definition                          | In-patient setting |
|--------------|----------------------------------|------------------------------------------------------------------|------------------------------|--------------------------------------------------------|-------------------|
| Bejan 2013   | Retrospective time series single center | USA, Washington University of Washington                        | 100                          | NR                                                     | ICU               |
| Kumamaru 2016| Retrospective case series single center | USA, Massachusetts Brigham and Women's Hospital                   | 125                          | acute pulmonary embolism                               | NR                |
| Bodduluri 2013| Retrospective case-control multi-center (national data) | USA, Iowa The University of Iowa                                  | 153                          | smokers with or without COPD and non-smokers           | NR                |
| Biesiada 2014| Prospective case series single center | USA, Cincinnati Children’s Hospital Medical Center & University of Cincinnati | 347                          | current tonsillitis, adenotonsillar hypertrophy or obstructive sleep apnea | Surgery           |
| Reamaroon 2018| Retrospective time series single-center | USA, Michigan University of Michigan                                | 401                          | mild hypoxia and acute hypoxic respiratory failure     | NR                |
| Vinson 2015  | Retrospective case series multi-center | USA, California the Kaisers Permanente CREST Network              | 593                          | acute pulmonary embolism                               | ED                |
| Huesch 2018  | Retrospective case series single center | USA, Pennsylvania Milton S. Hershey Medical Center                 | 1133                         | individuals suspected of pulmonary embolism           | ED                |
| Mortazavi 2017| Retrospective time series single center | USA, Connecticut Yale University                                 | 5214                         | patients undergoing cardiovascular procedures: CABG, PCI and ICD procedures | Surgery           |
| Pham 2014    | Retrospective case series single center | France CHU de Caen, Caen & Hôtel Européen Georges-Pompidou, Paris | NR (100)                     | individuals suspected of having Venous thromboembolism | NR                |
| Rochefort 2015| Retrospective time series single center | Canada, Quebec McGill University                                 | 1649 (2000)                  | individuals suspected of having Venous thromboembolism | various           |
| Silva 2017   | Prospective before-after multi-center (3 centers) | France University Teaching Hospital of Purpan, Toulouse; Hospital Dieu Hospital, Narbonne; Saint Eloi Hospital, Montpellier | 136                          | hemodynamic instability, respiratory failure, multiple trauma, nontraumatic coma, and postoperative complication of abdominal surgery | ICU               |
| Gonzalez 2018| Prospective time series multi-center, multinational | USA Binham and Women’s Hospital (on behalf of the COPD and ECLIPSE Study investigators) | 11655                        | smokers with or without COPD                           | various           |
| Tian 2017    | Retrospective case series single center | Canada, Quebec Mcgill University                                 | 2819 (4000)                  | individuals suspected of having Venous thromboembolism | various           |
| Study       | Study Design                  | Countries and institution(s)                                                                 | Number of patients (records) | Population/disease definition                          | In-patient setting |
|-------------|-------------------------------|---------------------------------------------------------------------------------------------|------------------------------|-------------------------------------------------------|--------------------|
| Choi 2018   | Prospective case series       | Mayo Clinic, Scottsdale; National Jewish Health, Denver; University of Washington Medical Center, Seattle & Veracyte Inc. | 139 (403)                   | suspected interstitial lung disease                   | NR                 |
| Yu 2014     | Retrospective case series     | Massachusetts Brigham, and Women’s Hospital & Harvard Medical School,                         | NR (10,330)                 | individuals suspected of pulmonary embolism           | NR                 |
| Swartz 2017 | Retrospective case series     | USA, New York University & Mount Sinai St. Luke’s Hospital                                  | NR (2400)                   | individuals suspected of having Venous thromboembolism| various            |
| Liu 2013    | Retrospective case series     | California Kaiser Permanente                                                               | NR (2466)                   |                                                        | ICU                |
| Haug 2013   | Retrospective case series     | USA, Utah LDS Hospital and Intermountain Medical Centre                                   | NR (362,924)                |                                                        | ED                 |
| Dublin 2013 | Retrospective case series     | USA, Seattle Group Health Research Institute & University of Washington                     | NR (5000)                   |                                                        | NR                 |
| Phillips 2014| Prospective case series      | UK, Llanelli Swansea University, Aberystwyth University & Hywel Dda University Health Board | 181                         | with and without COPD                                 | various            |
| Hu 2016     | Retrospective case series     | USA, Minnesota University of Minnesota                                                     | NR (8909)                   |                                                        | Surgery            |
| Jones 2018  | Retrospective case/time series| USA, Utah & Washington VA Salt Lake City Health Care System, University of Utah & George Washington University | NR (911)                    | individuals suspected of pneumonia                   | ED                 |

NA: Not applicable. NR: Not reported. USA: United States of America. COPD: Chronic obstructive pulmonary disease. ECLIPSE: Evaluations of COPD Longitudinally to Identify Predictive Surrogate Endpoints. UK: United Kingdom. CAGB: Coronary artery bypass grafting. PCI: Percutaneous coronary intervention. ICD: Implantable cardioverter defibrillator. ICU: Intensive care unit. ED: Emergency department.

Algorithm performance. It was also the single preferred outcome in three studies. About half of the studies additionally reported sensitivity, specificity, and accuracy. One study reported specificity with sensitivity set at 90% and 95% to ensure that few disease positive cases were missed. The single study that developed a CAD system measured the ROC AUC and model accuracy.

**Infection or sepsis**

The search yielded 2659 hits. Screening the titles and abstracts lead to 2562 being excluded. The full texts of the remaining 97 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=41), population (n=4); intervention (n=6) and outcomes (n=14).

A total of 31 studies were finally included in this systematic literature review. Four of these were ongoing trials. The study selection process is depicted in Figure 3.

**Study characteristics.** Of the included studies, 24 were conducted in the US. Three studies were conducted outside the US; one in France; one in the Netherlands and one in the UK. In total, 21 studies were retrospective and six were prospective. There were 21 single-center studies, and six multi-center studies. Seven studies were time series and 18 studies were case series, one was a case-control and one was a matched-controlled study.
Table 7. Overview of the algorithms developed to detect respiratory distress or failure.

| Study               | Predicted disease | NLP assertion classification | symbolic classifiers | rule or probability based | KNN | ONYX | RF | LR, LASSO penalized | LR, LASSO regularization | LR, not specified gradient (descent) boosting | Maximum Entropy | SVM | Partial least-squares regression | NeGEV | hierarchical classification | Bayesian network | neural network | J48 | JRIP | PART |
|---------------------|-------------------|------------------------------|----------------------|--------------------------|-----|------|----|---------------------|--------------------------|-------------------------|-----------------|-----|---------------------------------|--------|-----------------------------|-----------------|----------------|-----|-----|-----|
| Reamaroon 2018      | ARDS              |                              |                      |                          |     |      |    | ✓                  | ✓                       | ✓                       | ✓               |     |                                 |        |                             |                 |                |     |     |     |
| Gonzalez 2018       | COPD, ARDE        |                              |                      |                          |     |      |    | ✓                  | ✓                       | ✓                       | ✓               |     |                                 |        |                             |                 |                |     |     |     |
| Bodduluri 2013      | COPD              |                              |                      | ✓                        |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Phillips 2014       | COPD              |                              |                      | ✓                        |     |      |    |                    |                          | ✓                       | ✓               | ✓   |                                 |        |                             |                 |                |     |     |     |
| Bejan 2013          | Pneumonia         | ✓                            |                      |                          |     |      |    | ✓                  | ✓                       | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Dublin 2013         | Pneumonia         | ✓                            |                      | ✓                        |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Haug 2013           | Pneumonia         | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Hu 2016             | Pneumonia         | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Liu 2013            | Pneumonia         | ✓                            |                      | ✓                        |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Choi 2018           | Pneumonia         | ✓                            |                      | ✓                        | ✓   |      |    | ✓                  | ✓                       | ✓                       | ✓               | ✓   |                                 |        |                             |                 |                |     |     |     |
| Jones 2018          | Pneumonia         | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Silva 2017          | Postintubation distress |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Mortazavi 2017      | Postoperative respiratory failure |                              |                      |                          |     |      |    |                    |                          | ✓                       | ✓              | ✓   |                                 |        |                             |                 |                |     |     |     |
| Vinson 2015         | Pulmonary embolism | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Yu 2014             | Pulmonary embolism | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Huesch 2018         | Pulmonary embolism | ✓                            |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Kumamaru 2016       | Pulmonary embolism* |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Pham 2014           | Pulmonary embolism, DVT |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Rochefort 2015      | Pulmonary embolism, DVT |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Swartz 2017         | Pulmonary embolism, DVT |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Tian 2017           | Pulmonary embolism, DVT |                              |                      |                          |     |      |    |                    |                          | ✓                       |                |     |                                 |        |                             |                 |                |     |     |     |
| Biesiada 2014       | Respiratory depression |                            |                      |                          |     |      |    |                    |                          | ✓                       | ✓              |     |                                 |        |                             |                 |                |     |     |     |

*A computer aided detection system was developed for measuring the right ventricular/left ventricular axial diameter ratio and detecting pulmonary embolism. ARDS: Acute respiratory distress syndrome. ARDE: Acute respiratory disease events. COPD: Chronic obstructive pulmonary disease. DVT: Deep vein thrombosis.
| Study            | Algorithm | Sensitivity | Specificity | NPV  | PPV  | negative LR | positive LR | Accuracy | Prevalence | OR   | RR   | ROC AUC | Diagnostic yield |
|------------------|-----------|-------------|-------------|------|------|-------------|-------------|----------|------------|------|------|--------|------------------|
| Kumamaru 2016    | CAD       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        |            |      |      |        |                  |
| Bodduluri 2013   | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        |            |      |      |        |                  |
| Hu 2016          | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        |            |      |      |        |                  |
| Mortazavi 2017   | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        |            |      |      |        |                  |
| Rochefort 2015   | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Silva 2017       | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Vinson 2015      | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Biesiada 2014    | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Choi 2018        | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Gonzalez 2018    | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Phillips 2014    | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Reamaroon 2018   | ML        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Bejan 2013       | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Dublin 2013      | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Haug 2013        | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Liu 2013         | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Pham 2014        | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Swartz 2017      | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Tian 2017        | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Yu 2014          | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Huesch 2018      | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |
| Jones 2018       | NLP       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          |      |      |        |                  |

NLP: Natural language processing. ML: Machine learning. CAD: Computer aided detection. NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio. RR: Risk ratio. ROC AUC: Receiver operating characteristic area under the curve.

The smallest studies included patients with leukemia and combat casualty patients. Four studies had a sample size below 1000, three had a sample size between 1001–10,000 and 12 had a sample size larger than 10,000. Eight studies had samples even larger than 50,000.

Majority of the published studies evaluated data from different wards; several studies included patients admitted only to the ICU and surgical ward, less often the General ward and Emergency Department. Of these, 23 studies included data collected at their own hospital; and four utilized previously collated databases.

The characteristics of all published studies are given in Table 9.

CDS systems. The machine learning algorithms evaluated in the studies were developed to predict a range of diseases. These included sepsis, acute kidney injury, surgical site infections, central line-associated...
Figure 3. Study selection - infection or sepsis. Pop. = Population.
### Table 9. Design aspects of published studies on infection or sepsis.

| Study           | Study Design                        | Country and institution(s)                                                                 | Number of patients (records) | Population/disease definition | In-patient setting |
|-----------------|-------------------------------------|-------------------------------------------------------------------------------------------|------------------------------|-------------------------------|-------------------|
| Ahmed 2015      | Retrospective case series single center | USA, Minnesota Mayo Clinic Rochester                                                      | 944                          | NR                            | ICU               |
| Brasier, 2015   | Prospective case series multi-center (3 sites) | USA, Texas Aspergillus Technology Consortium & University of Texas                        | 57                           | Leukemia                      | NR                |
| Dente, 2017     | Prospective case series single center | USA, Maryland Emory University, Walter Reed National Military Medical Centre              | 73                           | Combat casualty patients      | NR                |
| Hu, 2016        | Retrospective case series single center | USA, Minnesota University of Minnesota                                                   | NR (8,909)                   | NR                            | General           |
| Konerman, 2017  | Retrospective time series single center | USA, Michigan University of Michigan                                                      | 1,233                        | Chronic hepatitis c           | NR                |
| Legrand, 2013   | Prospective case series single center | France, Paris Hôpital Européen Georges Pompidou Assistance Publique-Hopitaux de Paris     | 202                          | Infective endocarditis        | Surgery           |
| Mani, 2014      | Retrospective case series single center | USA, New Mexico University of New Mexico                                                 | 299                          | Sepsis                        | ICU               |
| Mao 2018        | Retrospective case series multi-center (5 centers) | USA University of California, Stanford Medical Centre, Oroville Hospital, Bakersfield Heart Hospital, Cape Regional Medical Centre, Beth Israel Deaconess Medical Center | 359,390                      | NR                            | various           |
| Sanger, 2016    | Prospective time series single center | USA, Washington University of Washington                                                | 851                          | Open-abdominal surgery patients | Surgery           |
| Scicluna, 2017  | Prospective case series multi-center (2 sites + national database) | Netherlands & UK Amsterdam Academic Medical Center, Utrecht University Medical Center & UK Genomic Advances in Sepsis study | 787                          | Sepsis                        | ICU               |
| Sohn, 2016      | Retrospective case series single center | USA, Minnesota Mayo Clinic Rochester                                                     | 751                          | Colorectal surgery patients   | Surgery           |
| Taylor, 2018    | Retrospective case series single center | USA, Connecticut Yale University School of Medicine,                                       | 55,365 (80,387)              | Suspected urine tract infection | ED                |
| Hernandez 2017  | Retrospective case series single center | UK, London Imperial College Healthcare NHS Trust                                          | > 500,000                    | NR                            | NR                |
| Study           | Study Design                          | Country and institution(s)                                      | Number of patients (records) | Population/disease definition                                                                 | In-patient setting |
|-----------------|---------------------------------------|----------------------------------------------------------------|----------------------------|----------------------------------------------------------------------------------------------|-------------------|
| Bartz-Kurycki 2018 | Retrospective case series multi-center (national database) | USA, Texas University of Texas                                  | 13,589                      |                                                                                               | Surgery           |
| Beeler 2018      | Retrospective case-control single center | USA, Indiana University Health Academic Health Center              | NR (70,218)                 | Central venous line with or without central line-associated bloodstream infections             | NR                |
| Bihorac 2018     | Retrospective time series single center | USA, Florida University of Florida Health                        | 51,457                      |                                                                                               | Surgery           |
| Chen 2018        | Retrospective matched pairs (1:1 case matching) single center | USA, Kansas University of Kansas Health System                   | 358                         | Stage 3 AKI and non-AKI controls                                                              | NR                |
| Cheng 2017       | Retrospective case series single center | USA, Kansas University of Kansas Medical Center                  | 33,703 (48,955)             |                                                                                               | NR                |
| Desautels 2016   | Retrospective case series single center | USA, California Dascena Inc.& University of California            | NR (21,176)                 |                                                                                               | ICU               |
| Koyner 2015      | Retrospective time series single center | USA, Chicago University of Chicago                                | NR (121,158)                |                                                                                               | NR                |
| LaBarbera 2015   | Retrospective case series single center | USA, Pennsylvania Pinnacle Health Hospital, Harrisburg            | 198                         | Clostridium difficile infection                                                               | NR                |
| Mohamadlou 2018  | Retrospective time series multi-center (2 sites) | USA, Dascena Inc., University of California & Stanford University | 68,319                      |                                                                                               | ICU               |
| Nemati 2018      | Retrospective time series multi-center (3 sites) | USA, Georgia Emory University School of Medicine & Georgia Institute of Technology | 69,938                      |                                                                                               | ICU               |
| Parreco 2018     | Retrospective time series single center | USA, Florida University of Miami                                  | NA (22,201)                 |                                                                                               | ICU               |
| Taneja 2017      | Prospective case series single center | USA, Illinois University of Illinois                             | 444                         | Suspected sepsis                                                                             | NR                |
| Weller 2018      | Retrospective case series single center | USA, Minnesota Mayo Clinic Rochester                              | 1,283                       | Colorectal surgery patients                                                                   | Surgery           |
| Wiens 2014       | Retrospective case series single center | USA single center not specified                                  | NR (69,568)                 |                                                                                               | various           |

NA: Not applicable. NR: Not reported. USA: United States of America. UK: United Kingdom. ICU: Intensive care unit. ED: Emergency department. AKI: Acute kidney injury.
bloodstream infections\textsuperscript{77,86}, \textit{Clostridium difficile}\textsuperscript{33,88}, pulmonary \textit{aspergillosis}\textsuperscript{89}, bacteremia\textsuperscript{90}, fibrosis\textsuperscript{91}, urine tract infection\textsuperscript{35,74} and infections in general\textsuperscript{19}.

Almost half of the studies compared different machine learning algorithms, while the others focused only on Bayesian algorithms\textsuperscript{13,92}, decision tree algorithms\textsuperscript{64}, ensemble algorithms\textsuperscript{85,71,82,83,90,93}, regression algorithms\textsuperscript{33,78,85}, regularization algorithms\textsuperscript{83,88} and rule learning\textsuperscript{15}. The most frequently applied model was random forest (15 studies) followed by logistic regression (10 studies), support vector machines (5 studies), naive Bayes (5 studies) and gradient tree boosting (5 studies).

One study compared three different sampling methods for handling class imbalance; under-sampling the majority class (RANDu), over-sampling the minority class (RANDo) and synthetic minority over-sampling (SMOTE). This was a very large study including more than 500,000 patients to predict the onset of infections\textsuperscript{39}. The authors found that SMOTE outperformed the other techniques and improved model sensitivity. Two other very large studies used the RANDu method\textsuperscript{80} and mini-batch stochastic gradient descent with backpropagation\textsuperscript{92}. No other studies were concerned with imbalance in disease positive and negative classification.

Machine learning models were trained and validated in 26 studies and subsequently tested in an independent dataset in four studies\textsuperscript{35,72,76,77}.

The machine learning algorithms used are illustrated in Table 10.

\textbf{Outcome measures.} The most frequently reported outcome measure was the ROC AUC. Three studies did not report this measure: Ahmed \textit{et al.} 2015\textsuperscript{70} developed an algorithm based on decision rules; Legrand \textit{et al.} 2013\textsuperscript{32} was primarily interested in identifying risk factors of AKI after cardiac surgery; and Scicluna \textit{et al.} 2017\textsuperscript{73} was primarily concerned with identifying genetic biomarkers of sepsis.

Sensitivity and specificity were reported together in 14 studies\textsuperscript{35,70–72,74,75,78,81–84,87,90,97}. When specificity was not reported, sensitivity was reported together with PPV; and when sensitivity was not reported, this was due to sensitivity being set at a fixed value to report other diagnostic performance measures. In relation to the prior observation, more studies reported PPV than NPV. Four studies reporting likelihood ratios reported both negative and positive likelihood ratios\textsuperscript{80,74,41,84}.

An overview of measured outcomes is illustrated in Table 11.

\textbf{Ongoing studies.} Four trials are currently ongoing, one in Germany and the others in the USA, all concerned with the prediction of sepsis. Three of them are prospective studies and one is retrospective. The retrospective study aims to develop a prediction algorithm based on claims data, EHRs, risk factors and survey data of an estimated 50,000 adult patients admitted to the ED. The German study NCT03661450\textsuperscript{95} is a single-arm trial evaluating the utility of a CDS system to identify SIRS or sepsis from EHRs in a pediatric ICU population. Another single-arm trial NCT03655626\textsuperscript{95} is concerned with implementing a sepsis prediction algorithm in clinical practice as an early warning system. NCT03644940\textsuperscript{95} is comparing two versions of InSight introduced into clinical practice as an early warning system.

\textbf{Discussion and conclusions} This systematic literature review shows that over the last 2 decades, there has been an increased interest in CDS as means of supporting clinicians in acute care. CDS has been investigated for several applications ranging from the detection of health conditions\textsuperscript{76,81}, to the prediction of deterioration or adverse events\textsuperscript{80,55,76,83,90}. Applications also include therapy guidance, as well as updating clinicians on new or changed recommendations\textsuperscript{96}. CDS can also provide guidance by predicting clinical trajectories for different patient profiles over time\textsuperscript{97}.

From rule-based algorithms and simple regression models, CDS has evolved to encompass a multitude of techniques in Machine-Learning\textsuperscript{98}. These techniques can be dependent on the problem selected and the data types used. Across the three disease areas investigated, the frequent use of random forest classifiers (28.1%), support vector machines (21.9%), boosting techniques (20.3%), LASSO regression (18.8%) and unspecified logistic regression models (10.9%) were observed. The use of more complex modeling such as maximum entropy, Hidden Markov Models (for temporal data analysis) as well as Convolutional Neural Networks has also emerged over the last few years. In the respiratory distress area, the use of NLP models is more common as radiology reports and clinical notes are the main source of input. Different image analysis techniques have been developed to aid in the prediction and diagnosis of respiratory events from radiology images.

Typical measures of NLP model performance include sensitivity, specificity and predictive values. In measuring ML algorithm performance, sensitivity, specificity and ROC AUC are more common. A wide range of outcome measure were reported in research on less-investigated health conditions\textsuperscript{40,62}; and also when uncommon, more complex algorithms were compared to basic algorithms\textsuperscript{72,74,81,84}. This is not surprising given the novelty of these applications.

Many of the ML algorithms and all of the NLP models covered in this work were based on medical data collected in certain clinical sites rather than publicly available data. Datasets from national audits, completed studies or other online sources can additionally play a role, particularly in model validation and testing. This could aid in the adoption and wider use of CDS systems. In this SLR, publicly available datasets were mainly utilized for developing prediction models of heart arrhythmias\textsuperscript{29–31}, hypotension\textsuperscript{2}, septic shock\textsuperscript{26,33,40,41}, COPD\textsuperscript{32}, pneumonia\textsuperscript{33} and a range of infections\textsuperscript{33,76,78,81,84,86}. In only three cases were they used for testing model performance in sepsis and septic shock prediction; this included the Insight algorithm\textsuperscript{35,85,93}.

Most of the studies identified in this SLR were retrospective and originated in the USA where electronic health records (EHR)
## Table 10. Overview of machine learning algorithms evaluated in studies on infection or sepsis.

| Study                  | Predicted disease | Machine learning algorithm |
|------------------------|-------------------|-----------------------------|
| Ahmed 2015             | AKI               | ✓                           |
| Legrand, 2013          | AKI               | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cheng 2017             | AKI               | ✓                           |
| Koyner 2015            | AKI               | ✓                           |
| Bihorac 2018           | AKI, sepsis       |                             |
| Mohamadiou 2018        | AKI, Stage 2/3    | ✓                           |
| Chen 2018              | AKI, Stage 3      | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Dente, 2017            | bacteremia        |                             |
| Beeler 2018            | CLABSI            | ✓                           |
| Pareco 2018            | CLABSI            | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| LaBarbera 2015         | clostridium difficile | ✓                           |
| Wiens 2014             | clostridium difficile | ✓                           |
| Konerman, 2017         | fibrosis          |                             |
| Hernandez 2017         | infection         | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Brasier, 2015          | pulmonary aspergilosis | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Mani, 2014             | sepsis            | ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Mao, 2018              | sepsis            | ✓                           |
| Study            | Predicted disease | Rule learning | NB | AODE | lazy Bayesian rules | Bayesian GLM | Bayesian network analysis | CART | decision tree classifier | neural network | RF | (extreme) gradient boosting | adaptive boosting | ensemble classifier | k nearest neighbor | MARS | GPS | Lasso penalized LR | LR, not specified | SVM | generalized additive model | GLM | stepwise regression | polynomial linear model | polytomous spline regression | Weibull PH model | L2-regularized LR | elastic net regularization |
|------------------|-------------------|---------------|----|------|---------------------|--------------|--------------------------|------|-------------------------|----------------|----|------------------------|-------------------|-------------------|------------------|------|-----|------------------|------------------|-----|--------------------|-----------------|------------------------|------------------------|------------------------|-----------------|--------------------|--------------------------|
| Scicluna, 2017   | sepsis            | ✓             |    |      |                     |              |                          |      |                         |                 |    |                        |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Desautels 2016   | sepsis            |               |    |      |                     |              |                          |      |                         |                 |    |                        |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Nemati 2018      | sepsis            |               |    |      |                     |              |                          |      | ✓                       |                 |    |                        |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Taneja 2017      | sepsis            | ✓             | ✓  | ✓    |                     |              |                          |      |                         |                 |    | ✓                       |                   |                   |                  |      |     | ✓                |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Sanger, 2016     | SSI               | ✓             | ✓  | ✓    |                     |              |                          |      | ✓                       |                 |    | ✓                       |                   |                   |                  |      |     | ✓                |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Sohn, 2016       | SSI               |               |    |      |                     |              |                          |      |                         |                 |    | ✓                       |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Bartz-Kurycki 2018 | SSI              |               |    |      |                     |              |                          |      | ✓                       |                 |    | ✓                       |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Weller 2018      | SSI               | ✓             | ✓  | ✓    |                     |              |                          |      | ✓                       |                 |    | ✓                       |                   |                   |                  |      |     | ✓                |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Hu 2016          | SSI, UTI, pneumonia, sepsis |               |    |      |                     |              |                          |      |                         |                 |    | ✓                       |                   |                   |                  |      |     |                   |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |
| Taylor, 2018     | UTI               | ✓             | ✓  | ✓    | ✓                   | ✓             | ✓                        | ✓    | ✓                       |                 |    | ✓                       |                   |                   |                  |      |     | ✓                |                  |     |                     |                 |                        |                         |                       |                 |                   |                          |

AKI: Acute kidney injury. SSI: Surgical site infection. UTI: Urinary tract infections. CLABSI: Central line-associated bloodstream infections. NB: Naive Bayes. AODE: Averaged one dependence estimators. CART: Classification and regression tree. RF: Random forest. MARS: Multivariate Adaptive Regression Splines. GPS: Generalized path seeker algorithm. LR: Logistic regression. SVM: Support vector machine. GLM: Generalized linear model. PH: Proportional hazards.
Table 11. Overview of measured outcomes in studies predicting sepsis or infection.

| Study             | Sensitivity | Specificity | NPV  | PPV  | negative LR | positive LR | Accuracy | Prevalence | OR  | RR  | ROC AUC |
|-------------------|-------------|-------------|------|------|-------------|-------------|----------|------------|-----|-----|---------|
| Ahmed 2015        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Brasier, 2015     | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Dente, 2017       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Hu, 2016          | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Konerman, 2017    | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Legrand, 2013     | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Mani, 2014        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Mao 2018          | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Sanger, 2016      | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Scicluna, 2017    | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Sohn, 2016        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Taylor, 2018      | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Hernandez 2017    | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Bartz-Kurycki 2018| ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Beeler 2018       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Bihorac 2018      | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Chen 2018         | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Cheng 2017        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Desautels 2016    | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Koyner 2015       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| LaBarbera 2015    | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Mohamadiou 2018   | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Nemati 2018       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Parreco 2018      | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Taneja 2017       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Weller 2018       | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |
| Wiens 2014        | ✓           | ✓           | ✓    | ✓    | ✓           | ✓           | ✓        | ✓          | ✓   | ✓   | ✓       |

NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio. RR: Risks ratio. ROC AUC: Receiver operator curve area under the curve.

are commonly used. This makes it easier to access and compile large amounts of patient-level information. Many of the studies on shock and infection/sepsis based their models on data extracted from EHRs and utilized large sample sizes. The diversity in the identified CDS systems makes it challenging to draw conclusions on methodology. The lack of comparisons between different classifiers within studies, especially for the indication of shock, adds to this challenge. To assess the effectiveness of ML algorithms, future research should evaluate multiple algorithms on standard well-labeled datasets.

Class imbalance can be an important issue when training classifiers on datasets for the conditions highlighted in this work. Unequal distributions can arise naturally between disease negative
and positive classes when forming validation sets, particularly when disease prevalence is low\textsuperscript{85}. We refer the reader to several machine learning reviews that have addressed this issue\textsuperscript{96–101}. Another important issue in forming disease positive classes relates to the analysis of repeated-measures within subjects, for example, when clinical records are available for each hospitalization day. Several studies have approached this by selecting the first record indicating positive for a health condition. Few researchers have utilized all records and corrected for within-subject variation. An example is the selection of cases depending on observed correlation decay\textsuperscript{82}.

In all three areas investigated, the number of retrospective studies exceeded by far the number of prospective studies conducted in a clinical setting. This highlights the challenges in substantiating clinical performance while bringing new clinical decision tools to routine in-hospital patient care. Examples of algorithms that can be integrated in clinical practice include InSight\textsuperscript{19,46} and Sepsis Watch\textsuperscript{47} which are intended for predicting sepsis and septic shock.

The current systematic literature review did not search multiple bibliographic databases or clinical trial registers; and focused on diagnostic performance rather than other outcomes. In fact, during study screening, trials that evaluated the impact of early warning systems on measures of clinical workflow, rate of re-admissions and/or mortality were discarded as they are somehow out of the focus of this work. This implies that there may be more CDS systems used in practice for the three populations investigated within this research, where the outcomes measured are different. Limiting the search to publications in English and to studies conducted in particular countries; and the exclusion of study protocols identified from the bibliographic database search without checking for later publications from the same authors may have further limited the studies selected. Nevertheless, studies identified within each population represented a diverse range of models applied in different hospital settings trained to predict a range of health conditions. The most widely researched conditions were sepsis and septic shock, venous thromboembolisms, acute kidney injury and surgical site infections.

Specific challenges were identified in collecting sufficient data for training CDS systems on hemodynamic instability. Patients who are, for example, at risk of hemorrhage due to a traumatic injury need to be carefully monitored; and the speed by which they reach a critical state may influence data and study management. It may also be difficult to find healthy volunteers who are willing to undergo procedures like lower body negative pressure which can be unpleasant\textsuperscript{86}. Identification of cases in need of hemodynamic interventions can lend towards larger sample size\textsuperscript{89}. Other conditions that need further attention are clostridium difficile and CLABSIs. Prediction models were driven by almost perfect specificity and very low (<10%) sensitivity\textsuperscript{77,83,86,88}. Considering that these studies used a wide range of features from the EHRs and a large number of patients, except LaBarbera, Nikiforov\textsuperscript{83}, there is a need to better understand the risk factors to improve sensitivity.

Based on the literature reviewed in this work, as well as several recent surveys and workshops, we would recommend the following points to be addressed when bringing a new CDS tool to critical care\textsuperscript{19,44–46,104}:

- Integrating CDS in clinical workflows without adding unnecessary extra work to busy clinical teams. The CDS101 toolbox by HIMMS highlights the “CDS five rights”, which are certainly applicable to critical care\textsuperscript{105}.
- Providing the right information in the right intervention format, to the right person at the right point in their workflow, and through the right channel.
- Developing tools and concrete proof-points able to assess CDS efficacy in the clinic. This also highlights the importance of providing continuous feedback to clinicians.
- The importance of easy to use user interfaces and focusing on human-computer interaction during deployment.
- Efficient training that is available when needed.
- Being aware of alert or alarm fatigue and not overloading clinicians with alerts due to CDS. The intensive care unit is already plagued with alarms, and if anything, CDS should help in reducing alarms by bundling alerts according to underlying conditions.
- Displaying the rationale for decisions as well as the underlying data to clinical users would lead to improved adoption.
- Understanding ethical challenges for CDS, as well as a careful risk assessment in every site before deployment\textsuperscript{106}.
- Being able to repeat/standardize implementation across organizations – most prospective studies reviewed in this work covered single centers. Only a few were multi-center studies.

**Data availability**

**Underlying data**

All data underlying the results are available as part of the article and no additional source data are required.

**Extended data**

Figsphere: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 1-Search for shock (hemodynamic (in-stability) in MEDLINE. docx. https://doi.org/10.6084/m9.figshare.9892109.v1\textsuperscript{25}.

Figsphere: Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended
data - Table 2-Search strategy for respiratory distress or respiratory failure in MEDLINE.docx. https://doi.org/10.6084/m9.figshare.9892112.v1#

Figshare: Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 3-Search strategy for infection or sepsis in MEDLINE.docx. https://doi.org/10.6084/m9.figshare.9892115.v1#

Reporting guidelines
Figshare: PRISMA checklist for ‘Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review’. https://doi.org/10.6084/m9.figshare.9894107.v1#

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Acknowledgments
We would like to thank Mark Connolly from Global Market Access Solutions Sàrl for his contribution during the whole project.

References

1. Molina JA, Seow E, Heng BH, et al.: Outcomes of direct and indirect medical intensive care unit admissions from the emergency department of an acute care hospital: a retrospective cohort study. BMJ Open. 2014; 4(11): e005553. PubMed Abstract | Publisher Full Text | Free Full Text
2. Winters B, Custer J, Galvagno SM Jr, et al.: Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual Saf. 2012; 21(11): 894–902. PubMed Abstract | Publisher Full Text
3. Rothschild JM, Landrigan CP, Cronin JW, et al.: The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in critical care. Crit Care Med. 2005; 33(8): 1694–700. PubMed Abstract | Publisher Full Text
4. Donovan JL, Kanaan AO, Thomson MS, et al.: Effect of clinical decision support on psychotropic medication prescribing in the long-term care setting. J Am Geriatr Soc. 2010; 58(5): 1005–7. PubMed Abstract | Publisher Full Text
5. Field TS, Rochon P, Lee M, et al.: Computerized clinical decision support during medication ordering for long-term care residents with renal insufficiency. J Am Med Inform Assoc. 2009; 16(4): 480–5. PubMed Abstract | Publisher Full Text | Free Full Text
6. Kennedy CC, Campbell G, Garg AX, et al.: Piloting a renal drug alert system for prescribing to residents in long-term care. J Am Geriatr Soc. 2011; 59(9): 1757–9. PubMed Abstract | Publisher Full Text | Free Full Text
7. Tamblyn R, Eguale T, Buckeridge DL, et al.: The effectiveness of a new generation of computerized drug alerts in reducing the risk of injury from drug side effects: a cluster randomized trial. J Am Med Inform Assoc. 2012; 19(4): 635–43. PubMed Abstract | Publisher Full Text | Free Full Text
8. Marasinghe KM: Computerised clinical decision support systems to improve medication safety in long-term care homes: a systematic review. BMJ Open. 2015; 5(5): e006339. PubMed Abstract | Publisher Full Text | Free Full Text
9. Quinn CC, Clough SS, Minor JM, et al.: WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther. 2008; 10(3): 160–8. PubMed Abstract | Publisher Full Text | Free Full Text
10. Correa E, Lau AY, Tsafnat G, et al.: The changing nature of clinical decision support systems: a focus on consumers, genomics, public health and decision safety. Yearb Med Inform 2009; 84–95. PubMed Abstract | Publisher Full Text | Free Full Text
11. Agoritsas T, Heen AF, Brandt L, et al.: Decision aids that really promote shared decision making: the pace quickens. BMJ. 2015; 350: g7624. PubMed Abstract | Publisher Full Text | Free Full Text
12. Vincent JL, Elain S, Pearse R, et al.: Improving detection of patient deterioration in the general hospital ward environment. Eur J Anaesthesiol. 2018; 35(5): 325–333. PubMed Abstract | Free Full Text
13. Cox JC, Sadinaj Y, Schnier KE, et al.: Higher Quality and Lower Cost from Improving Hospital Discharge Decision Making. J Econ Behav Organ. 2016; 131(8): 1–16. PubMed Abstract | Publisher Full Text | Free Full Text
14. Toheng JE, Bakken S, Bates DW, et al.: Optimizing Strategies for Clinical Decision Support. In: The Learning Health System Series. N.A.o. Medicine. Editor. Washington DC USA. 2017. Reference Source
15. Duncan H, Hutchison J, Parshuram CS: The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. J Crit Care. 2006; 21(3): 271–8. PubMed Abstract | Publisher Full Text
16. Parshuram C, Duncan HP, Joffe AR, et al.: Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children. Crit Care. 2011; 15(4): R184. PubMed Abstract | Publisher Full Text | Free Full Text
17. Chapman SM, Bray J, Oulton K, et al.: “The Score Matters”: wide variations in predictive performance of 18 paediatric track and trigger systems. Arch Dis Child. 2017; 102(6): 487–496. PubMed Abstract | Publisher Full Text | Free Full Text
18. Phillips. [Accessed 2nd July 2018]. 2018. Reference Source
19. Potes C, Corroy B, Xu-Wilson M, et al.: A clinical prediction model to identify patients at high risk of hemodynamic instability in the pediatric intensive care unit. Crit Care. 2017; 21(1): 282. PubMed Abstract | Publisher Full Text | Free Full Text
20. Hravnak M, Devita MA, Clontz A, et al.: WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther. 2008; 10(3): 160–8. PubMed Abstract | Publisher Full Text | Free Full Text
21. Gaieski DF, Mikkelsen ME, Band RA, et al.: Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. Crit Care Med. 2010; 38(4): 1045–53. PubMed Abstract | Publisher Full Text | Free Full Text
22. Critical care statistics. [cited 2019 September 10]. Reference Source
23. Mayr FB, Yende S, Angels DC: Epidemiology of severe sepsis. Virology. 2014; 5(1): 4–11. PubMed Abstract | Publisher Full Text | Free Full Text
24. Sakr Y, Jaschinski U, Witeboole X, et al.: Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit. Open Forum Infect Dis. 2018; 5(12): ojy313. PubMed Abstract | Publisher Full Text | Free Full Text
25. Medic G, Kleiśl MK, Atalahl L, et al.: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. https://doi.org/10.6084/m9.figshare.9894107.v1#
26. Medic G, Kleiśl MK, Atalahl L, et al.: Working title: Evidence-based Clinical
Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 2-Search strategy for respiratory distress or respiratory failure in MEDLINE. docx. figshare. Dataset. 2019. http://www.figshare.com/articles/36849/tf.9892112.v1

27. Medic G, Kiellöf MK, Atallah L, et al.: Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 3-Search strategy for infection or sepsis in MEDLINE.docx. figshare. Dataset. 2019. http://www.figshare.com/articles/10.6084/m9.figshare.9892115.v1

28. Ghasi D, Li J, Cori L, et al.: Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. J Biol Inform. 2017; 46: 19–31. Published Abstract | Publisher Full Text

29. Li Q, Rajagopalan C, Clifford GD: Ventricular fibrillation and tachycardia classification using a machine learning approach. IEEE Trans Biomed Eng. 2014; 61(6): 1607–13. Published Abstract | Publisher Full Text

30. Ebrahimzadeh E, Kalantari M, Joulani M, et al.: Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal. Comput Programs Biomed. 2018: 165: 53–67. PubMed Abstract | Publisher Full Text

31. Strothoff N, Strothoff C: Detecting and interpreting myocardial infarction using fully convolutional neural networks. Physiol Meas. 2018; 40(1): 015001. PubMed Abstract | Publisher Full Text

32. Donald R, Howells T, Piper I, et al.: Forewarning of hypotensive events using a Bayesian artificial neural network in neurocritical care. J Clin Monit Comput. 2018; 33(1): 39–51. PubMed Abstract | Publisher Full Text

33. Hu Z, Melton GB, Moeller ND, et al.: Accelerating Chart Review Using Automated Methods on Electronic Health Record Data for Postoperative Complications. AMIA Annu Symp Proc 2016; 2016: 1822–1831. PubMed Abstract | Free Full Text

34. Mahajan D, Dong Y, Saxon LA, et al.: Performance of an automatic arrhythmia classification algorithm: comparison to the ALTITUDE electrocardiography panel adjudications. Pacing Clin Electrophysiol. 2014; 37(7): 889–99. PubMed Abstract | Publisher Full Text

35. Mao G, 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(1): e017833. PubMed Abstract | Publisher Full Text | Free Full Text

36. Rejini N, Zimmer G, Malaya Y, et al.: Using support vector machines on photoplethysmographic signals to discriminate between hypovolemia and euvolemia. PLos One. 2018; 13(3): e0196087. PubMed Abstract | Publisher Full Text | Free Full Text

37. Šídersis G, Pourhomayoun M, Kalantarian H, et al.: A flexible data-driven commodity feature extraction framework. Comput Biol Med. 2016; 73: 165–72. PubMed Abstract | Publisher Full Text | Free Full Text

38. Becker S, Katz SD, Horowitz LL, et al.: Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. JAMA Cardiol. 2016; 1(9): 1014–1020. PubMed Abstract | Publisher Full Text | Free Full Text

39. Becker S, Howells T, Horowitz LL, et al.: Early Identification of Patients With Acute Decompensated Heart Failure. J Card Fail. 2018; 24(6): 367–362. PubMed Abstract | Publisher Full Text | Free Full Text

40. Calvert J, Desautels T, Chetlalally U, et al.: High-performance detection and early prediction of septic shock for alcohol-use disorder patients. Am Med Surg (Lon). 2016; 8: 50–5. PubMed Abstract | Publisher Full Text | Free Full Text

41. Henry K, Hager DN, Pronovost PJ, et al.: A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med. 2016; 8(398): 398ra122. PubMed Abstract | Publisher Full Text | Free Full Text

42. Panahiazar M, Taslimitehrani V, Pereira N, et al.: Using EHRs and Machine Learning for Heart Failure Survival Analysis. Stud Health Technol Inform. 2015; 216: 40–4. PubMed Abstract | Free Full Text

43. NCT02954971: Optimized Multi-modality Machine Learning Approach During Cardio-toxic Chemotherapy to Predict Airing Heart Failure (MERMAID). Reference Source

44. NCT03582651: Measurement of Hemodynamic Responses to Lower Body Negative Pressure (LBNP). Reference Source

45. NCT03235193: Predictive algorithm for Evaluation and Intervention in SEpsis (PREVISE). Reference Source

46. NCT03644940: Subpopulation-Specific Sepsis Identification Using Machine Learning. Reference Source

47. NCT03656626: Implementation and Evaluations of Sepsis Watch. Reference Source

48. Bejan CA, Vanderswende L, Evans HL, et al.: On-time clinical phenotype prediction based on narrative reports. AMIA Annu Symp Proc. 2013: 103–10. PubMed Abstract | Free Full Text

49. Kumamaru KK, George E, Aghayev A, et al.: Implementation and Performance of Automated Software for Cognitive-to-Left Ventricular Diameter Ratio From Computed Tomography Pulmonary Angiography Images. J Comput Assist Tomogr. 2016; 40(3): 387–92. PubMed Abstract | Publisher Full Text | Free Full Text

50. Bondook S, Newell JD, Jr, Hoffman EA, et al.: Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. Acad Radiol. 2013; 20(5): 527–36. PubMed Abstract | Publisher Full Text | Free Full Text

51. Biesiada J, Chidambaran V, Wagner M, et al.: Genetic risk signatures of opioid-induced respiratory depression following pediatric tonsillectomy. Pharmacogenomics. 2014; 15(4): 1749–1762. PubMed Abstract | Publisher Full Text | Free Full Text

52. Reamaron N, Sjödín MW, Lin K, et al.: Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome. IEEE J Biomed Health Inform. 2019; 23(1): 407–415. PubMed Abstract | Publisher Full Text | Free Full Text

53. Vinson DR, Morley JE, Huang J, et al.: The Accuracy of an Electronic Pulmonary Embolism Severity Index Auto-Populated from the Electronic Health Record: Setting the stage for computerized clinical decision support. Appl Clin Inform. 2015; 6(2): 318–33. PubMed Abstract | Publisher Full Text | Free Full Text

54. Huesch MD, Chenien R, Labin S, et al.: Evaluating Report Text Variation and Informativeness: Natural Language Processing of CT Chest Imaging for Pulmonary Embolism, J Am Coll Radiol. 2018; 15(3 Pt B): 554–562. PubMed Abstract | Publisher Full Text

55. Marzavi BJ, Desai N, Zhang J, et al.: Prediction of Adverse Events in Patients Undergoing Major Cardiac Procedures, IEEE J Biomed Health Inform. 2017; 21(6): 1719–1729. PubMed Abstract | Publisher Full Text

56. Gonzalez G, Ash SY, Vegas-Sánchez-Ferrero G, et al.: Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. Am J Respir Crit Care Med. 2018; 197(2): 193–203. PubMed Abstract | Publisher Full Text

57. Choi Y, Liu TT, Pankratz DG, et al.: Identification of usual interstitial pneumonia pattern using RNA-Seq and machine learning: challenges and solutions. BMC Genomics. 2016; 17(Suppl 2): 101. PubMed Abstract | Publisher Full Text | Free Full Text

58. Yu S, Kumamaru KK, George E, et al.: Classification of CT pulmonary angiography reports by presence, chronicity, and location of pulmonary embolism with natural language processing. J Biomed Inform. 2014; 52: 386–93. PubMed Abstract | Publisher Full Text | Free Full Text

59. Swartz J, Koziatek C, Theobald J, et al.: Creation of a simple natural language processing tool to support an imaging utilization quality dashboard. Int J Med Inform. 2017; 101: 93–99. PubMed Abstract | Publisher Full Text | Free Full Text

60. Liu V, Clark MP, Mendosa M, et al.: Automated identification of pneumonia in chest radiograph reports in critically ill patients. BMC Med Inform Decis Mak. 2013; 13: 50. PubMed Abstract | Publisher Full Text | Free Full Text

61. Haug PJ, Ferraro JP, Holmen J, et al.: An ontology-driven, diagnostic modeling system. J Am Med Inform Assoc. 2013; 20(1): e102. PubMed Abstract | Publisher Full Text | Free Full Text

62. Dublin S, Baldwin E, Walker RL, et al.: Natural Language Processing to identify pneumonia from radiology reports. Pharmacoepidemiol Drug Saf. 2013; 22(8): 834–41. PubMed Abstract | Publisher Full Text | Free Full Text

63. Jones BE, South BR, Shao Y, et al.: Development and Validation of a Natural Language Processing Tool to Identify Patients Treated for Pneumonia across VA Emergency Departments. Am J Med Inform. 2018; 25(1): 122–128. PubMed Abstract | Publisher Full Text | Free Full Text

64. Rochefort CM, Verma AD, Eguale T, et al.: A novel method of adverse event detection can accurately identify venous thromboembolism (VTEs) from narrative electronic health record data. J Am Med Inform Assoc. 2015; 22(1): 155–65. PubMed Abstract | Publisher Full Text | Free Full Text

65. Pham AD, Névéol A, Lavergne T, et al.: Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. BMC Biomatics. 2014; 15: 266. PubMed Abstract | Publisher Full Text | Free Full Text

66. Silva S, Ait Aissa D, Cocquet P, et al.: Combined Thoracic Ultrasound Assessment during aSuccessful Weaning Trial Predicts Postextubation Distress. Anesthesiology. 2017; 127(4): 666–674. PubMed Abstract | Publisher Full Text | Free Full Text

67. Phillips C, Mac Parthalain N, Syed V, et al.: Short-Term Intra-Subject Variation in Exhaled Volatile Organic Compounds (VOCs) in COPD Patients and Healthy
Open Peer Review

Current Peer Review Status: ✓ ✓

Version 2

Reviewer Report 04 December 2019
https://doi.org/10.5256/f1000research.23644.r57156

© 2019 Nikolakopoulos S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ Stavros Nikolakopoulos
Department of Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

My comments were adequately addressed in the text. No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 28 November 2019
https://doi.org/10.5256/f1000research.23644.r57155

© 2019 Kovacevic M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ Milena Kovacevic
Department of Pharmacokinetics and Clinical Pharmacy, Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia

Thank you for addressing my comments. I have no further comments to make.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Pharmacokinetics and Clinical Pharmacy; Patient outcomes.
The review summarizes the utilization of clinical decision support (CDS) systems in three selected states in critical care – shock/hemodynamic (in-)stability; respiratory distress/failure; and infection/sepsis. The background of the study has a strong rationale.

The study comprised the results from primary sources, describing models/algorithms used to detect and alert clinicians to the presence of these conditions, as well as models/algorithms developed to predict deterioration in an individual patient state, leading to these selected conditions.

The systematic review was performed and the findings are presented in line with the PRISMA guidelines. Variables for which data were sought were clearly stated (PICOS) in Table 1.

Specific comments:
- What I found especially beneficial for the readers and future research in this area, is Table 2 with the presented collected data used for training algorithms.
- It would be beneficial to provide additional information whether an internal or external validation was performed - within Table 4 (measured outcomes in studies on shock), Table 8 (measured outcomes in studies on respiratory distress/failure) and Table 11 (measured outcomes in studies on infection/sepsis).
- What was the rationale for including the studies predicting acute kidney injury within the Infection/sepsis results section? If it is about the decline in glomerular filtration rate due to hypotension seen in sepsis, it might have been presented within the Shock section.
- Table 7: include the abbreviations for ARDS (Acute respiratory distress syndrome), ARDE (Acute respiratory disease events) and DVT (deep vein thrombosis) below the Table.
Table 9: include the abbreviation for AKI (Acute kidney injury) below the Table.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Not applicable

Are the conclusions drawn adequately supported by the results presented in the review?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Pharmacokinetics and Clinical Pharmacy; Patient outcomes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
(or more extended) time frame, which seems slightly inconsistent. Some discussion concerning this choice would be enlightening.

- There seems to be some confusion with terminology, with unknown consequences on the review's results. The authors seem to separate "machine learning" methods, from "statistical" methods (Table 1: "Multivariable hierarchal logistic regression models*** (models which are based only on statistics - but there is no machine learning)", as an exclusion criterion). This is clearly not the suitable platform to resolve this issue, but, the distinction between machine learning and statistics is not at all that clear. Specifically, under the term "supervised learning", any regression method (statistics) could be classified. So, logistic regression IS a machine learning method. So is LASSO and several other methods reported. Again, this is not the appropriate place for going into further details, but there is certainly some confusion, especially when in the results Logistic regression keeps appearing as a preferred method.

- Again concerning terminology, the term "accuracy" appears often in the results section. Sometimes it is reported as a different outcome than i.e. ROC AUC, sensitivity and specificity. All the latter methods are quantifying "accuracy" in some way and some clarification is needed.

**Minor comments:**
- Table 1: Treatment/Intervention, a parenthesis is missing.
- Tables 7 & 10: Maybe reverse the orientation of the column titles, it is impossible to read on a screen.

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**
Yes

**Are sufficient details of the methods and analysis provided to allow replication by others?**
Yes

**Is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are the conclusions drawn adequately supported by the results presented in the review?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Statistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com