UNIVERSITY OF AMSTERDAM

# UvA-DARE (Digital Academic Repository)

## The secrets of sepsis

*A data-driven approach to improve its diagnostic work-up and treatment*

Schinkel, M.

**Publication date**
2023

**Document Version**
Final published version

Link to publication

**Citation for published version (APA):**
Schinkel, M. (2023). *The secrets of sepsis: A data-driven approach to improve its diagnostic work-up and treatment*. [Thesis, fully internal, Universiteit van Amsterdam].

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

Download date:28 Nov 2023

# THE SECRETS OF SEPSIS

## A DATA-DRIVEN APPROACH TO IMPROVE ITS DIAGNOSTIC WORK-UP AND TREATMENT

MICHIEL SCHINKEL

# THE SECRETS OF SEPSIS:

## A data-driven approach to improve its diagnostic work-up and treatment

Michiel Schinkel

ISBN: 978-94-93278-43-1

Layout and cover: Off Page, Amsterdam

# THE SECRETS OF SEPSIS:

## A data-driven approach to improve its diagnostic work-up and treatment

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 26 mei 2023, te 16.00 uur

door

Michiel Schinkel
geboren te Utrecht

## PROMOTIECOMMISSIE

| | | |
|---|---|---|
| Promotores: | prof. dr. W.J. Wiersinga | AMC-UvA |
| | prof. dr. K.D.P.W.B. Nanayakkara | Vrije Universiteit Amsterdam |
| Copromotores: | dr. F. Holleman | AMC-UvA |
| Overige leden: | dr. J.J. Hoogerwerf | Radboudumc |
| | prof. dr. T. van der Poll | AMC-UvA |
| | prof. dr. A.P.J. Vlaar | AMC-UvA |
| | prof. dr. H.F. Lingsma | Erasmus Universiteit Rotterdam |
| | prof. dr. J.M. Prins | AMC-UvA |
| | dr. P.W.G. Elbers | VU Medisch Centrum |

Faculteit der Geneeskunde

# TABLE OF CONTENTS

# CHAPTER 1

## GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

*On an ordinary Wednesday in February 2016, the definition of sepsis changed. Based on consensus among international experts, the then-applicable definition was revised to the new Sepsis-3. Some patients who fulfilled the previous Sepsis-2 criteria did not meet those for Sepsis-3 and vice versa. It begged the question, who really has sepsis? I was about to start my last year of medical school when I read the new sepsis definition for the first time. I remember feeling puzzled that the sepsis syndrome I had seen and learned about was suddenly obsolete. In the weeks thereafter, I started reading up on the sepsis literature. To my surprise, I found that there were almost as many opinions as there were experts. Various articles outlined arguments for and against the new definition. On top of that, evidence contradicted the benefits of even the most conventional treatment strategies. No sepsis-specific therapies existed despite decades of research. I realized that sepsis still held many secrets, and as chance would have it, I spent the next four years trying to unravel some of these.*

## THE PAST AND PRESENT OF SEPSIS

The sepsis syndrome can be traced back as early as 3.000 BC when physicians in Egypt described systemic inflammation following traumatic injuries[1]. The term sepsis, derived from the word "sepo" [σηπω], which means "I rot," was first coined in a poem by Homer in ancient Greece around 750 BC[2]. Over time, sepsis was deemed a systemic infectious disease resulting from invading pathogens spreading through the bloodstream, often described as "blood poisoning"[3]. Considering this theory, the discovery of antibiotics and their potential to eradicate invading organisms should have been the cure for sepsis. However, many sepsis patients still died despite eliminating the pathogen. It was thus postulated that sepsis should be regarded as an interaction between the microbe and the host[4]. In the years following this paradigm shift, an expert panel reached a consensus about the first sepsis definition in 1992[5]. Through further revisions in 2003 and 2016, we have come to the third international consensus definition[6,7]. Sepsis is now defined as a life-threatening syndrome in which the body's response to microbial invasion is dysregulated and causes organs to dysfunction[7]. The host's immune response, characterized by concomitant hyperinflammation and immunosuppression, derails and fails to return to homeostasis[8]. Based on the Sepsis-3 definition, it is estimated that 49 million patients are affected globally, with 11 million sepsis-related deaths each year (Figure 1)[9].

## THE DIAGNOSIS OF SEPSIS

Many factors contribute to the high morbidity and mortality of sepsis. Arguably, the most important is the challenge of diagnosing sepsis early and accurately[1]. The presenting symptoms, such as fever, and laboratory results, such as leukocytosis, are nonspecific. Sepsis can originate in many sites throughout the body, and the complex interaction between the host and microbes can be disrupted in multiple ways. Classical cases, such as meningococcal sepsis, are rare, and even those diagnoses may not be evident in the early stages[1].

The current sepsis definition helps address the primary challenge of risk stratification, identifying those most likely to deteriorate while avoiding excessive resource use in low-risk patients[1]. Time pressure hinders an extensive diagnostic workup and confirmation of infection in sepsis[11]. Awaiting


**Figure 1. An overview of the worldwide impact and consequences of sepsis.** (*adapted from sepsisamsterdam.nl[10]).

culture results before initiating treatment may have disastrous consequences since patients can deteriorate or die before the results are known. Consequently, the Sepsis-3 criteria do not require a proven infection to diagnose sepsis, and the only measurable feature of the sepsis definition is organ dysfunction[7]. The Sequential Organ Failure Assessment (SOFA) score is used to measure the severity of organ dysfunction, which can also be roughly estimated based on mental status, systolic blood pressure, and respiratory rate, constituting the quickSOFA (qSOFA)[7]. A significant concern is that once multiple organs become dysfunctional, the impact of potentially beneficial treatment strategies that target the causative pathogen or the host response can be limited[1]. To impact the disease course meaningfully, we may need to intervene sooner. Consequently, there has been increasing interest in diagnosing sepsis earlier using biomarkers, electronic health record surveillance, and computational approaches, as well as raising awareness among clinicians and the lay public[1].

## THE MANAGEMENT OF SEPSIS

Several historical discoveries, such as the antiseptic effects of handwashing by Ignaz Semmelweis (1847) and antibiotics by Alexander Fleming (1928), have helped lower sepsis mortality rates[12,13]. The current management strategies can broadly be categorized into three groups: resuscitation and supportive care, infection control, and modulating the host response (Figure 2)[14]. The sepsis-induced organ dysfunction can be alleviated by resuscitating the fluid balance and oxygen status. Both (surgical) interventions to remove infected tissue ('ubi pus, ibi evacua') and the administration of antimicrobial therapy can help prevent the further spreading of the causative microbes. Yet, we lack effective sepsis-specific interventions targeting the host response[14].

Landmark papers evaluating sepsis treatment strategies at the beginning of the 21st century showed promising results. Some were even recommended by the international Surviving Sepsis Campaign, which has provided evidence-based guidelines for sepsis management since the early 2000s[16]. A paper by Rivers et al. on Early Goal-Directed Therapy (EGDT) assessed the value of standardized operating procedures to retain physiological targets such as a central venous oxygen

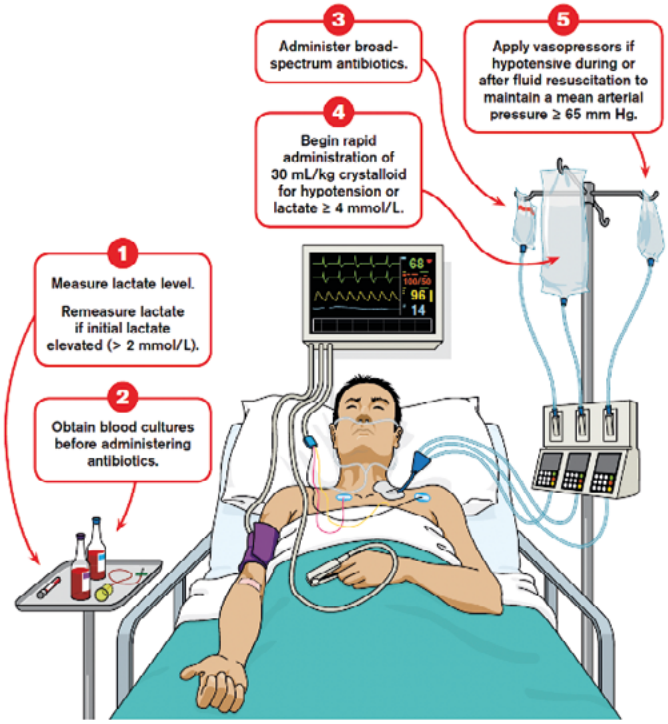



**Figure 2. An overview of general management strategies for sepsis in the Surviving Sepsis Campaign guidelines.** (*adapted from the surviving sepsis campaign bundle infographic[15])

saturation over 70%[17]. Patients treated according to these targets were more likely to survive (69.5%) than those who were not (53.5%). Unfortunately, three subsequent, large multicenter trials and a meta-analysis could not find such benefits[18–21]. A similar pattern was observed for human activated protein C (drotrecogin alpha activated). This host response modulating drug, with antithrombotic, anti-inflammatory, and profibrinolytic properties, seemed to decrease sepsis mortality by 6.1% in the PROWESS trial but did not do so in follow-up studies[22–24]. Another treatment strategy that has received significant attention is the early administration of antibiotics when sepsis is first suspected. A highly cited paper from 2006 showed that every hour of delay in antibiotic therapy decreased the chance of survival[25]. Although these findings were based on retrospective data, sepsis guidelines since have been guided by the results of this study. However, subsequent meta-analyses and the single randomized controlled trial (RCT) on this subject have not shown clear benefits of early antibiotic treatment for suspected sepsis[26–28]. The recommendations of the Surviving Sepsis Campaign for the abovementioned treatments have been altered, weakened, or removed entirely[29]. Controversial, nonreplicable, and disappointingly negative trial results have since dominated the search for effective management strategies for sepsis[30]. It seems clear that we need a different approach to unravel the secrets of sepsis.

## CHALLENGES IN SEPSIS RESEARCH

One of the most pressing challenges in sepsis research is the heterogeneity in the study populations[30]. Sepsis is an "umbrella term" for patients with infections at various sites, with different microorganisms, comorbidities, and genetic predisposition[14,31]. It is unlikely that a magic bullet exists which is effective in all these patients. In recent years, sepsis researchers have shifted their focus to stratifying patients into more homogenous subgroups, which may respond more similarly to certain treatment strategies[32,33]. Early attempts to create homogeneous clusters of sepsis patients have used biomarkers, genome-wide blood gene expression profiles, and clinical data in combination with advanced analytical approaches[34–36]. Identifying these subgroups, which are usually called endotypes or phenotypes (see table 1 for definitions), may help predictively enrich study populations and test new therapies only in sepsis patients with similar characteristics and pathobiological profiles that indicate the patient may benefit from a certain intervention[37]. An example of this approach is a study by Shakoory et al. that reanalyzed the effect of interleukin-1 antagonist Anakinra in sepsis only in those who had features of Macrophage Activation Syndrome (MAS)[38]. While the original study in all sepsis patients was stopped early because of futility, Anakinra use was associated with improved survival in those with concurrent features of MAS[38,39]. Although we are now gradually moving from a "one-size-fits-all" to a "few-sizes-fit-most" approach, there may be potential to use novel technologies to help understand sepsis on a deeper level and move towards truly personalized and optimized sepsis care.

## DATA AND THE AGE OF ARTIFICIAL INTELLIGENCE

The start of the Information Age in the mid-20$^{th}$ century heralded the widespread use of data in medicine[40]. In the early days, small datasets were used to gain general insights into specific disease processes. Data-driven approaches gained momentum in the late 1990s and early 2000s when the introduction of electronic health record systems and monitoring devices facilitated large-scale medical data capture[41]. In the following years, physicians were progressively inundated with data, and novel technologies were needed to sort through all that data and generate insights. Artificial Intelligence (AI), the scientific discipline that aims to emulate human thought processes and decision-making through computer programs, is such a technology [42]. AI's fundamental value proposition is the ability to analyze and learn from large amounts of data and detect patterns of information[42]. With the current transition into the Age of AI, it is hoped that these tools will be increasingly able to act on those data patterns and positively augment human decision-making.

To appreciate the value of AI for clinical purposes, we need to understand what AI is. Rather than being a single technology, AI is a collection of fields, such as robotics, computer vision, natural language processing, and machine learning[43]. Machine learning aims to let the computer recognize patterns from examples instead of teaching them through explicit instructions. It has been the most tangible manifestation of AI in the healthcare industry[44]. We can further subdivide machine learning into supervised, unsupervised, and reinforcement learning. With supervised learning, the goal is to train computer models to recognize the associations between inputs (predictors) and outputs (labels)[44]. When these associations have been learned, they can be used to predict future outcomes

**Table 1.** Definition of specialized terms used throughout this thesis. *adapted from chapter 35 of the Sepsis Codex: Artificial Intelligence in Sepsis[45].

| Term | Definition |
|---|---|
| **Artificial Intelligence (AI)** | The scientific discipline that aims to design and understand computer systems that mimic human cognition. AI has many subfields such as robotics, computer vision, and machine learning. |
| **Machine Learning** | Machine learning is the subfield of AI that uses data and algorithms to emulate human learning. Popular techniques within the machine learning domain are supervised learning, unsupervised learning and reinforcement learning. |
| **Endotypes** | Endotypes are subgroups within a condition that are distinguished by pathobiological mechanisms. |
| **Phenotypes** | Phenotypes are subgroups within a condition that are distinguished by observable traits. |

with new but similar input data. In unsupervised learning, there is no need for labels. The main goal is to identify specific structures and closely related data points[44]. They can, for example, be used to cluster together patients with similar traits. Finally, reinforcement learning works in situations with sequential actions. Through trial and error, the computer will try to learn the optimal series of steps to reach a desirable outcome[44].

## A DATA-DRIVEN APPROACH TO SEPSIS

Thinking back on the challenges in the sepsis research field, there is a clear need for more personalized sepsis management[30,37,46]. AI, and machine learning specifically, offers some solutions to this problem. Unsupervised methods can help find meaningful subtypes of patients with similar characteristics who may respond more similarly to treatments. When used effectively, this information may help enrich study populations and facilitate finding beneficial therapies. Furthermore, supervised machine learning techniques can support clinical decision-making by providing insights tailored to a specific patient. In the more distant future, the highly complex treatment of sepsis may be a great use case for reinforcement learning[47]. The management consists of multiple principles and actions, such as administering antibiotics and vasopressors, oxygen suppletion, and mechanical ventilation[29]. Combinations of these actions in different orders may yield different outcomes, and reinforcement learning may find the optimal sequence of actions to enhance the chances of full recovery for the individual patient.

The transition of AI into routine clinical care will have many technological, medical, and ethical challenges. Data privacy concerns, susceptibility to adversarial attacks, algorithmic bias, and liability issues only form the tip of the iceberg[48,49]. We are just beginning to understand the implications of decision-making augmented by algorithms. If we can sufficiently address these challenges and concerns, then data and machine learning can help solve many problems in sepsis research and management.

1

## AIMS AND OUTLINE OF THIS THESIS

This thesis investigates the cornerstones of sepsis management. We hypothesize that data and machine learning can optimize their use. The first chapters build on years of research and aim to create an evidence-based baseline of sepsis management, given the most recent clinical insights. The subsequent chapters aim to uncover specific aspects of sepsis care that may be improved through data-driven approaches while considering the challenges with AI in healthcare, as discussed above.

**Part 1** of the thesis focuses on improvement programs to optimize sepsis care. In **chapter 2**, we outline the value of the current sepsis management guidelines and the use of sepsis performance improvement programs to maximize compliance with those. Based on what we learned through this review, we implemented a sepsis performance improvement program, including a specialized sepsis response team, in our hospital and report the results in **chapter 3**.

In **part 2** of the thesis, we address the diagnostic workup of sepsis. In **chapter 4,** we conducted a narrative review to map the current artificially intelligent (AI) screening tools and decision support for sepsis. We also explore the potential for further improvements of such tools. **Chapter 5** presents the development, multicenter validation, and prospective evaluation of our AI tool to support the diagnostic workup of sepsis. We created a tool based on electronic health record data to predict the outcomes of blood cultures collected in the emergency department. The tool can help avoid blood culture testing in low-risk patients and prevent the high but often hidden costs of unnecessary blood culture testing. **Chapter 6** investigates potential barriers and facilitators to implementing our tool in clinical practice. Subsequently, we designed a randomized clinical trial to test the impact of implementing the blood culture prediction tool in daily practice, which will be discussed among the future perspectives in the general discussion.

**Part 3** of this thesis investigates aspects of the treatment of sepsis and ways in which data-driven approaches may help find effective therapies. Despite decades of research, searching for new sepsis therapies has yet to yield results. In contrast to the broader sepsis field, researchers discovered effective treatments for the coronavirus disease (COVID-19), a distinct subgroup of viral sepsis, within months after the first case of COVID-19 was described. **Chapter 7** outlines what sepsis researchers can learn from the COVID-19 pandemic. An essential difference between sepsis in general and COVID-19 specifically is that the COVID-19 diagnosis is much more homogeneous. In **chapter 8**, we further explore the value of using data-driven clustering techniques to decrease population heterogeneity in COVID-19. Building on these lessons regarding cohort heterogeneity, we aim to apply this knowledge to create homogeneous subgroups in sepsis to help find effective therapies. Currently, we still depend primarily on antibiotics to treat sepsis. The advised approach is to administer broad-spectrum antibiotics to all suspected sepsis patients as soon as possible. **Chapter 9** summarizes the (lack of) evidence for this practice. However, we still presume that some subgroups of sepsis patients benefit from early antibiotic treatment. In **chapter 10**, we apply clustering techniques to the sepsis population to find those subgroups.

The epilogue in **part 4** of this thesis starts with proposing a framework for implementing AI training in medical education in **chapter 11**. By establishing and growing data literacy in healthcare,

future physicians will be prepared to work with the tools we have created and implemented throughout this thesis. **Chapter 12** provides a general overview and discussion of the work in this thesis and the related literature. Finally, **chapter 13 and 14** summarize all the main results in both English and Dutch.

1

## REFERENCES


1. Angus, D. C. & Bindman, A. B. Achieving Diagnostic Excellence for Sepsis. *JAMA* **327**, 117–118 (2022).
2. Geroulanos, S., Douka, E. T., Geroulanos, S. & Douka, E. T. Historical perspective of the word "sepsis". *Intensive Care Medicine 2006 32:12* **32**, 2077–2077 (2006).
3. Finfer, S. R., Vincent, J.-L., Angus, D. C. & van der Poll, T. Sepsis NEJM 2013. *N Engl J Med* **9369**, 840–51 (2013).
4. Cerra, F. B. The systemic septic response: Multiple systems organ failure. *Crit Care Clin* **1**, 591–607 (1985).
5. Bone, R. C., Sibbald, W. J. & Sprung, C. L. The ACCP-SCCM consensus conference on sepsis and organ failure. *Chest* **101**, 1481–1483 (1992).
6. Levy, M. M. *et al*. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* **31**, 1250–1256 (2003).
7. Singer, M. *et al*. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 801 (2016).
8. van der Poll, T., van de Veerdonk, F. L., Scicluna, B. P. & Netea, M. G. The immunopathology of sepsis and potential therapeutic targets. *Nature Reviews Immunology* vol. 17 407–420 Preprint at https://doi.org/10.1038/nri.2017.36 (2017).
9. Rudd, K. E. *et al*. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
10. Home - Sepsis Amsterdam. https://sepsisamsterdam.nl/.
11. Angus, D. C. *et al*. A Framework for the Development and Interpretation of Different Sepsis Definitions and Clinical Criteria. *Crit Care Med* **44**, e113 (2016).
12. Noakes, T. D., Borresen, J., Hew-Butler, T., Lambert, M. I. & Jordaan, E. Semmelweis and the aetiology of puerperal sepsis 160 years on: an historical review. *Epidemiol Infect* **136**, 1 (2008).
13. Gaynes, R. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerg Infect Dis* **23**, 849 (2017).
14. Vincent, J.-L. Current sepsis therapeutics. *EBioMedicine* **0**, 104318 (2022).
15. SSC Adult Patients | SCCM. https://www.sccm.org/SurvivingSepsisCampaign/Guidelines/Adult-Patients.
16. Dellinger, R. P. The Future of Sepsis Performance Improvement. *Critical Care Medicine* vol. 43 1787–1789 Preprint at https://doi.org/10.1097/CCM.0000000000001231 (2015).
17. Rivers, E. *et al*. Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock. *New England Journal of Medicine* **345**, 1368–1377 (2001).
18. Mouncey, P. R. *et al*. Trial of Early, Goal-Directed Resuscitation for Septic Shock. *New England Journal of Medicine* **372**, 1301–1311 (2015).
19. The ProCESS Investigators. A Randomized Trial of Protocol-Based Care for Early Septic Shock. *New England Journal of Medicine* **370**, 1683–1693 (2014).
20. The ARISE Investigators and the ANZICS Clinical Trials Group. Goal-Directed Resuscitation for Patients with Early Septic Shock. *New England Journal of Medicine* **371**, 1496–1506 (2014).
21. Angus, D. C. *et al*. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. doi:10.1007/s00134-015-3822-1.
22. Bernard, G. R. *et al*. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* **344**, 699–709 (2001).
23. Lai, P. S. & Thompson, B. T. Why activated Protein C was not successful in severe sepsis and septic shock: Are we still tilting at windmills? *Curr Infect Dis Rep* **15**, 407–412 (2013).
24. Ranieri, V. M. *et al*. Drotrecogin Alfa (Activated) in Adults with Septic Shock. *New England Journal of Medicine* **366**, 2055–2064 (2012).
25. Kumar, A. *et al*. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med* **34**, 1589–1596 (2006).
26. Sterling, S. A., Miller, R., Pryor, J., Puskarich, M. A. & Jones, A. E. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis HHS Public Access. *Crit Care Med* **43**, 1907–1915 (2015).

27. Rothrock, S. G. *et al*. Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. *Annals of Emergency Medicine* vol. 0 Preprint at https://doi.org/10.1016/j.annemergmed.2020.04.042 (2020).

28. Alam, N. *et al*. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. *Lancet Respir Med* **6**, 40–50 (2018).

29. Evans, L. *et al*. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* (2021) doi:10.1007/s00134-021-06506-y.

30. Cohen, J. *et al*. Sepsis: A roadmap for future research. *The Lancet Infectious Diseases* Preprint at https://doi.org/10.1016/S1473-3099(15)70112-X (2015).

31. Vincent, J. L., van der Poll, T. & Marshall, J. C. The End of 'One Size Fits All' Sepsis Therapies: Toward an Individualized Approach. *Biomedicines 2022, Vol. 10, Page 2260* **10**, 2260 (2022).

32. Komorowski, M., Green, A., Tatham, K. C., Seymour, C. & Antcliffe, D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine* **0**, 104394 (2022).

33. Demerle, K., Angus, D. C. & Seymour, C. W. Precision Medicine for COVID-19: Phenotype Anarchy or Promise Realized? *JAMA - Journal of the American Medical Association* vol. 325 2041–2042 Preprint at https://doi.org/10.1001/jama.2021.5248 (2021).

34. Panacek, E. A. *et al*. Efficacy and safety of the monoclonal anti-tumor necrosis factor antibody F(ab') 2 fragment afelimomab in patients with severe sepsis and elevated interleukin-6 levels. *Crit Care Med* **32**, 2173–2182 (2004).

35. Scicluna, B. P. *et al*. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med* **5**, 816–826 (2017).

36. Seymour, C. W. *et al*. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* **321**, 2003 (2019).

37. Stanski, N. L. & Wong, H. R. Prognostic and predictive enrichment in sepsis. *Nature Reviews Nephrology* vol. 16 20–31 Preprint at https://doi.org/10.1038/s41581-019-0199-3 (2020).

38. Shakoory, B. *et al*. Interleukin-1 receptor blockade is associated with reduced mortality in sepsis patients with features of the macrophage activation syndrome: Re-analysis of a prior Phase III trial. *Crit Care Med* **44**, 275 (2016).

39. Opal, S. M. *et al*. Confirmatory interleukin-1 receptor antagonist trial in severe sepsis: a phase III, randomized, double-blind, placebo-controlled, multicenter trial. The Interleukin-1 Receptor Antagonist Sepsis Investigator Group. *Crit Care Med* **25**, 1115–1124 (1997).

40. Wartman, S. A. & Donald Combs, C. Medical education must move from the information age to the age of artificial intelligence. *Academic Medicine* **93**, 1107–1109 (2018).

41. Healthcare Data Growth: An Exponential Problem. https://www.nextech.com/blog/healthcare-data-growth-an-exponential-problem.

42. Schinkel, M., Paranjape, K., Nannan Panday, R. S., Skyttberg, N. & Nanayakkara, P. W. B. Clinical applications of artificial intelligence in sepsis: A narrative review. *Comput Biol Med* **115**, (2019).

43. Ohno-Machado, L., Nadkarni, P. & Johnson, K. Natural language processing: Algorithms and tools to extract computable information from EHRs and from the biomedical literature. *Journal of the American Medical Informatics Association* vol. 20 805 Preprint at https://doi.org/10.1136/amiajnl-2013-002214 (2013).

44. Panch, T., Szolovits, P. & Atun, R. Artificial intelligence, machine learning and health systems. *J Glob Health* **8**, (2018).

45. Borges, M., Hidalgo, J. & Perez-Fernandez, J. *The Sepsis Codex*. (Elsevier, 2022).

46. Maslove, D. M. *et al*. Redefining critical illness. *Nat Med* **28**, 1141–1148 (2022).

47. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* **24**, 1716–1720 (2018).

48. Kogetsu, A., Ogishima, S. & Kato, K. Authentication of patients and participants in health information exchange and consent for medical research: A key step for privacy protection, respect for autonomy, and Trustworthiness. *Front Genet* **9**, 167 (2018).

49. Finlayson, S. G., Chung, H. W., Kohane, I. S. & Beam, A. L. Adversarial Attacks Against Medical Deep Learning Systems. *Science (1979)* **363**, 1287–1289 (2018).

# PART I

## SEPSIS PERFORMANCE IMPROVEMENT PROGRAMS

# CHAPTER 2

## SEPSIS PERFORMANCE IMPROVEMENT PROGRAMS: FROM EVIDENCE TOWARD CLINICAL IMPLEMENTATION

Michiel Schinkel[1,2], Prabath W. B. Nanayakkara[2] and W. Joost Wiersinga[1,3*]

*Correspondence: w.j.wiersinga@amsterdamumc.nl
[1] Center for Experimental and Molecular Medicine, Amsterdam UMC, Location Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
Full list of author information is available at the end of the article

*Schinkel et al. Critical Care (2022) 26:77*
*https://doi.org/10.1186/s13054-022-03917-1*

## ABSTRACT

This article is one of ten reviews selected from the Annual Update in Intensive Care and Emergency Medicine 2022. Other selected articles can be found online at https://www.biomedcentral.com/collections/annualupdate2022.

Further information about the Annual Update in Intensive Care and Emergency Medicine is available from https://link. springer.com/bookseries/8901.

## INTRODUCTION

Since its launch in the early 2000s, the international Surviving Sepsis Campaign (SSC) has provided guidelines for the management of sepsis, most recently updated in 2021 [1]. The SSC aims to provide a standard of care for sepsis while increasing awareness among healthcare professionals and the general public. The goal is to reduce morbidity and mortality from sepsis and septic shock worldwide [2].

To facilitate the clinical implementation of the guidelines, the SSC bundles their recommendations into small groups of care processes that physicians should perform within a specific timeframe and that provides them with a concrete plan of action [1, 2]. Despite efforts to facilitate the successful implementation of the guidelines, adherence has been suboptimal, particularly regarding the microbiological work-up and administration of appropriate antibiotics [3]. Non-compliance to the SSC guidelines seems most prominent among emergency medicine and internal medicine physicians [4].

In response to the low adoption rates of (SSC) sepsis guidelines, individual hospitals and organizations have introduced sepsis performance improvement programs. Usually, dedicated physicians or research teams lead these initiatives and use screening tools, process changes in sepsis care pathways, and sepsis educational programs to optimize adherence to the standard of care [5]. The latest update of the SSC guidelines recommends that all hospitals and health systems have sepsis performance improvement programs [1].

In this chapter, we discuss the literature on the use and benefits of sepsis performance improvement programs to improve protocol adherence and provide practical insights for the clinical implementation of such programs in your hospital.

## DO 'ONE-SIZE-FITS-ALL' CARE BUNDLES IMPROVE SEPSIS OUTCOMES?

Sepsis performance improvement programs aim to improve adherence to a guideline or protocol for sepsis care, and they are almost exclusively studied in the context of the SSC care bundles [5]. When one aims to improve compliance rates to any guideline, one should first be convinced that this is a goal worth pursuing. In the case of the SSC guidelines, this debate has been ongoing for many years, and this paragraph presents only a brief overview of this reflective and meaningful discussion [6, 7].

Expert panelists on sepsis have created the SSC bundles, spearheaded by the Society of Critical Care Medicine (SCCM) and European Society of Intensive Care Medicine (ESICM), and endorsed by numerous medical societies [1, 6]. However, the evidence base for these bundles and the timeframes in which they should be performed have been a matter of debate [7–9]. One prominent example concerns adherence to early goal-directed therapy (EDGT), an early form of bundled care that was associated with significantly lower in-hospital mortality rates (30.5% vs. 46.5% in the usual care group) in a randomized study of 263 patients with sepsis or septic shock presenting to the emergency department of a tertiary hospital in the United States [10]. However, these results were not replicated in subsequent large randomized trials and meta-analyses [11–13]. Furthermore,

the value of individual bundle items, such as the 30 ml/kg fluid bolus and administration of antibiotics within 1 h to all patients, has been heavily debated because of conflicting results regarding the benefits [7, 14–16]. Moreover, fear exists that pressure to perform bundle items within a specific timeframe may promote harmful diagnostic tests and treatments, as was the case with the 2002 quality measure for the management of community-acquired pneumonia by the Centers for Medicare & Medicaid Services, which was later removed [9, 17].

Despite the limited evidence base that underlies some of the recommendations in the SSC bundles, the overall consensus, underscored by the endorsements from 35 international medical societies, seems to be that most of the care processes in the bundles will positively contribute to the management of the majority of sepsis patients [6]. Numerous observational studies have shown associations between improved bundle compliance and a reduction in mortality. An extensive 7.5-year study in 280 hospitals across Europe, South America, and the United States showed that overall mortality was significantly lower in high-compliance hospitals (29.0%) compared with low-compliance hospitals (38.6%) [18]. This study included 29,470 patients with sepsis or septic shock from emergency departments, regular wards, and intensive care units (ICUs) between January 1st 2005 and June 30th 2021. Notably, compliance was defined as high when sites completed the resuscitation bundle within 6 h for as few as 15% of their patients, suggesting that complete bundle adherence is only practical in a small subset of patients [18]. A similar project in Portugal studied the effects of adherence to the 6-h bundle in 897 patients with community-acquired sepsis in 17 ICUs [19]. Among those 897 patients, the core bundle was only completed within 6 h in 12% of the patients. The highest compliance was seen for the administration of vasopressors (78%) and the collection of cultures before antibiotic treatment (77%). In comparison, the lowest adherence was seen for blood culture collection in general (48%) and administration of antibiotics (52%) [19]. Compliance with the complete bundle was associated with decreased 28-day mortality, with an adjusted odds ratio (OR) of 0.44 (95% confidence interval [CI] 0.24–0.80) in sepsis and 0.49 (95% CI 0.25– 0.95) in septic shock. Other studies have found similar mortality benefits associated with improved SSC bundle adherence [20–22].

## SEPSIS IMPROVEMENT PROGRAMS: WHAT IS THE EVIDENCE?

Adherence to the SSC guidelines in hospitals and healthcare systems that have adopted them is still suboptimal [3, 5]. For example, a nationwide study in Finland showed complete guideline adherence in only 6 out of 92 ICU patients during the four-month study period, similar to rates found in other studies [5, 23]. Sepsis performance improvement programs may help improve compliance, and a 2015 systematic review and meta-analysis by Damiani and colleagues tried to quantify this effect [5]. The reviewers identified 50 observational studies with highly diverse improvement programs and study designs. Despite this heterogeneity, the meta-analysis showed that sepsis performance improvement programs were consistently associated with increased compliance with 6-h (OR 4.12, 95% CI 2.95–5.76) and 24-h (OR 2.57, 95%-CI 1.74–3.77) bundles and with reduced mortality (OR 0.66, 95%-CI 0.61–0.72). The mortality estimates are hard to interpret in this meta-analysis since they include in-hospital mortality as well as short- and long-term mortality.

Among the 50 studies included in the systematic review of Damiani et al., combinations of interventions using screening tools, process changes, and educational programs were independently associated with increased bundle compliance and reduced mortality [5]. It thus appears that having a sepsis performance improvement program in itself is more important than the specific content of the program. However, the best results were observed in programs with various simultaneous interventions for performance improvement and in hospitals where the initial compliance was lowest [5]. The following sections will discuss the most-studied interventions (implementation of sepsis screening tools, process changes in sepsis care pathways, and educational programs) and their effects in further detail.

## Sepsis Screening Tools

A primary focus of many performance improvement programs is using screening tools to identify sepsis early. Correct treatment can be initiated earlier if sepsis is recognized sooner, which is expected to improve patient outcomes [2]. Three randomized controlled trials (RCTs) have studied whether the use of screening tools can improve patient outcomes in sepsis [24–26]. Downing et al. used an electronic health record (EHR) alert to detect sepsis early in medical and surgical wards, based on modified sepsis criteria including laboratory results and vital signs [24]. However, the alert did not result in improved performance measures or patient outcomes.

Hooper and colleagues studied the effects of pager alerts whenever a patient in the medical ICU satisfied a modified version of the systemic inflammatory response syndrome (SIRS) criteria [25]. Again, the alerts did not result in any improved performance measures or decreased mortality rates. Only Shimabukuro and colleagues were able to show improvements in patient outcomes using automatically generated alerts in the EHR with their machine learning-based sepsis screening tool [26].

Among 142 patients in the US-based medical-surgical ICUs, the hospital length-of-stay (– 2.30 days), ICU length-of-stay (– 2.09 days), and in-hospital mortality (– 12.3%, absolute) were all significantly lower in the intervention group that used the automated sepsis screening tool [26]. One explanation for why this study was able to find beneficial effects is that it was the only one of the three to combine the alert with a mandatory and immediate evaluation of the patient to specifically address the potential diagnosis of sepsis, which can be regarded as an additional process change.

A problem in sepsis screening is that there is a plethora of different risk scores and screening tools which are currently used, such as the SIRS criteria, Modified Early Warning Score (MEWS), National Early Warning Score (NEWS), and quick Sequential Organ Failure Score (qSOFA). The accuracy of these risk scores is highly variable in the emergency department, regular wards, and the ICU [27].

Several extensive studies and reviews have evaluated which screening tool is most effective for suspected infection or sepsis [27–31]. The NEWS and MEWS consistently show a balance between sensitivity and specificity, both usually ranging between 0.40 and 0.80 [27, 29]. SIRS is more sensitive than specific, and qSOFA more specific than sensitive. None of these instruments seems superior to the others in identifying sepsis across studies [27–31]. The SSC guideline consequently does not

recommend using a particular tool [1]. Physicians should be aware of the benefits and limitations of the tools they use, and choices should be based on local preferences. The only exception is the use of qSOFA, which the guideline recommends against as a screening tool [1]. Although the qSOFA is highly specific, the poor sensitivity makes it unsuitable for screening purposes.

A limitation to all currently used tools is that they are susceptible to false positives because of the relatively low prevalence of sepsis, particularly in the general emergency department and ward populations [30]. Advanced computational approaches such as machine learning could provide a solution for this and may eventually replace the current, less complex risk scores. A systematic review and meta-analysis evaluating seven studies showed that machine learning algorithms outperform MEWS, SIRS, and qSOFA for sepsis prediction [32]. Additionally, monitoring through EHR systems with continuous data streams can detect sepsis even earlier than static risk scores. Van Wyk et al. showed this when their algorithm predicted sepsis onset in 377 ICU patients in the USA on average 205 min earlier than SIRS criteria would have [33]. However, many challenges still need to be overcome before safely introducing machine learning tools for sepsis into everyday clinical practice [34]. Some of these challenges were recently illustrated by the external validation of the Epic Sepsis Model, the machine learning-based screening tool for sepsis provided by the EHR vendor, Epic (Verona, WI, USA) [35]. This algorithm is widely adopted for sepsis screening, particularly in the USA. In a population of 2552 sepsis patients among 38,455 hospitalizations, the Epic Sepsis Model reached an area under the curve (AUC) of only 0.63 for sepsis recognition in an external validation [35]. Physicians using this tool evaluated an average of 109 patients based on sepsis screening alerts to detect only one case earlier than they would have without, putting a disproportionate burden on the healthcare system.

## Process Changes in Sepsis Care Pathways

Several studies have examined the effect of sepsis performance improvement programs using process changes to improve adherence to the SSC care bundles. After identifying a patient who may have sepsis, the diagnostic work-up and treatments should be promptly initiated. The most critical process change in sepsis care pathways studied in this regard is the implementation of sepsis (response) teams. Instead of putting the responsibility to act on a sepsis screening alert on one consulting physician, who may already care for multiple patients, dedicated teams are created to respond to sepsis alerts collectively. A prepost study by Viale et al. in Italian emergency departments showed that implementing a dedicated sepsis response team was associated with increased bundle adherence from 4.6 to 32%, improved appropriateness of the initial antibiotic therapy from 30 to 79%, and a hazard ratio of 0.64 (95% CI 0.43–0.94) for 14-day all-cause mortality [3]. In another study from Italy, these results were replicated in a multidisciplinary ICU [36]. In this setting, implementing a dedicated sepsis team was reported to be associated with a significant decrease in in-hospital mortality from 68 to 23%. Furthermore, the use of the dedicated sepsis team was significantly associated with decreased mortality in univariate logistic analysis (OR 0.28, 95% CI 0.10–0.79) [36]. However, the results of these studies should be interpreted cautiously, given their observational design and potential for confounding by indication.

Process changes other than implementing a dedicated sepsis team may also contribute to better bundle adherence when they improve the efficiency of the care workflow. Examples that have been extensively studied are printed or easily accessible protocols, standardized EHR order sets, daily auditing with weekly feedback, and nurse-driven sepsis protocols [5]. Nurse-driven sepsis protocols are a practical approach that acknowledges the essential role of nurses in the sepsis care pathways [37]. Their role is not formally described in the SSC guidelines, but they are often the first to triage patients and respond to their deteriorating condition. As an example, a Dutch study by Tromp et al. showed that a nurse-driven sepsis care bundle increased compliance with the complete bundle from 3.5 to 12.4% and the mean number of performed bundle elements within the appropriate timeframe from 3.0 to 4.2 [37]. Completion of four of the six individual bundle items, such as the measurement of serum lactate (23% to 80%) and the start of antibiotics within 3 h (38% to 56%), increased significantly. No significant changes in the in-hospital mortality rates or hospital length of stay were observed [37].

### Sepsis Educational Programs

Arguably, increased sepsis awareness is one of the primary reasons for better patient outcomes through SSC care bundle use. Therefore, education is an essential aspect of sepsis performance improvement programs, as it helps raise awareness among healthcare professionals. The 2015 systematic review about sepsis performance improvement programs by Damiani et al. included 17 studies in which only educational programs were used [5]. These included educational materials, lectures, bedside teaching, and simulation training, among others. Many of these education-only programs showed significantly increased bundle adherence and decreased mortality rates. An early observational cohort study in the USA by Nguyen et al. studied the effects of a comprehensive sepsis education program in a small cohort of 96 patients with sepsis in their ICU [38]. A mortality rate of 45% was observed when the compliance with SSC care bundles was high, but was 73% when SSC guidelines were largely disregarded ($p$ = 0.006). Another example of the effects of educational programs is the more extensive study by van Zanten and colleagues, which also reduced the limitations of the observational approach by using control groups and propensity score matching [22]. Implementation of educational programs in 52 participating hospitals was associated with an absolute increase of 23.6% in SSC bundle adherence and an absolute decrease in mortality rates of 5.8% in 8031 ICU patients with sepsis during the study period. No such associations were found in 8387 ICU patients in 30 non-participating hospitals over the same period.

## THE ROAD AHEAD

The discussion about the precise value of the SSC care bundles and the care processes within them will inevitably continue [6, 7]. Standardized expert care recommendations are indispensable for a syndrome with a mortality rate as high as it is in sepsis. However, such recommendations are often challenging to develop given the heterogeneity of sepsis and the weak and often contradicting evidence for its different treatment modalities [1, 13, 39]. Still, bundle adherence has consistently been associated with improved patient outcomes. An unanswered question is whether improved

patient outcomes are caused by the items in the care bundles, by increased awareness irrespective of bundle adherence, or whether they are just artifacts of confounding by indication. Well-controlled trials could potentially find a definitive answer to this question, further determining what matters most while implementing sepsis performance improvement programs. Such a trial will, however, be hard to carry out and needs sophisticated methodological design.

Sepsis improvement programs are associated with improved protocol compliance and can be helpful to improve protocol adherence when a hospital or healthcare system implements either the SSC sepsis guidelines or their version of a protocol for sepsis detection and treatment. Therefore, these programs should be used in any hospital with low adherence rates to local protocols. The program should ideally consist of various simultaneous interventions to promote bundle compliance optimally [5]. Those interventions can be sepsis screening tools, process changes in sepsis care pathways, and sepsis educational programs. However, the goal should never be to mandate 100% guideline adherence but to leave room to deviate from standardized protocols when appropriate.

In our university medical center, we initiated a sepsis performance improvement program in 2021. As an illustration, we provide the details about this program, including early lessons learned from the implementation process in Box 1. The flowchart for our sepsis response team set-up is visually presented in Fig. 1. A major takeaway is that the engagement of only a few clinical leaders per department seems insufficient in an emergency department's dynamic and continuous environment. Furthermore, the involvement of patient representatives is important when initiating a sepsis performance improvement program, as the values and perspectives of the main stakeholder should not be overlooked. In high-pressure situations, such as acute care for patients with suspected sepsis in the emergency department, treatment of the patient's physical state is prioritized over the mental state. However, systematically addressing important questions the patient may have could alleviate much of the mental stress they will likely experience. In Box 2, we summarize important questions to address from the viewpoint of a sepsis survivor who has been involved with our sepsis performance improvement program.

Finally, most studies investigating the benefits of bundled care and sepsis performance improvement programs used mortality reduction as an endpoint [5]. Already in 2005, an International Sepsis Forum (ISF) colloquium provided a broad set of outcome measures that sepsis studies can use beyond survival as the only and ultimate goal of sepsis care [40]. Nevertheless, the literature is still dominated by the pursuit of short-term survival benefits. During the coronavirus disease 2019 (COVID-19) pandemic, the ISF proposed an adjusted version of the original outcome set, which was adopted globally [40, 41]. Improving outcome parameters such as resource use, duration of invasive treatments, and the development of organ dysfunction that requires higher levels of care, suddenly became extremely valuable in a resource-scarce setting [42]. Future studies on sepsis performance iprovement programs and sepsis care bundles should similarly expand the core set of outcome measures to capture these additional benefits. In the era of shared decision-making and patient-centered care, we should acknowledge that there is more to life than death [43].

*Pre-implementation phase:*

- Retrospective and prospective evaluation of the current situation to identify opportunities for improvement. We noted:
  - » Sequential ED consultations by various specialists, which delayed appropriate care.
  - » Non-urgent triage codes in (elderly) patients with suspected sepsis.
- Involvement of patient representatives.

*Interventions:*

- Screening tool selected: MEWS (already in use and thus easy to incorporate).
- Process changes: Initiation of a sepsis response team, standardized notes and EHR order sets, daily audit and weekly feedback.
- Education: Launch of a dedicated website, pocket cards, talks at morning hand-over.

*Lessons learned so far:*

- Early challenges include behavior change and trust among all stakeholders that the new workflow will be efficient and may improve outcomes.
- The engagement of only a few clinical leaders per department seems insufficient for successful implementation, especially in the dynamic environment of an ED.

*ED* emergency department, *ICU* intensive care unit, *MEWS* Modified Early Warning Score, *EHR* electronic health record.

**Box 1. An example from the emergency department: creating a sepsis performance improvement program in a large university medical center.** The different phases of implementing a sepsis performance improvement program in the Amsterdam University Medical Center

- Acknowledge the signs that a patient is worried and take them seriously
- Communicate about the word "sepsis" and what it means
- Communicate the urgency that the potential sepsis is recognized
- Inform the patient about the use of a sepsis team or sepsis protocol
- Inform the patient about the plan of action, including possible tests, treatments, and other decisions to be made over the following hours
- Inform the patient about the effects/symptoms that can be expected from the treatment or progression of the syndrome

**Box 2. Essential aspects of emergency department sepsis care from the patient's point of view.** A summary of aspects to address during the evaluation of and conversation with a patient who may have sepsis

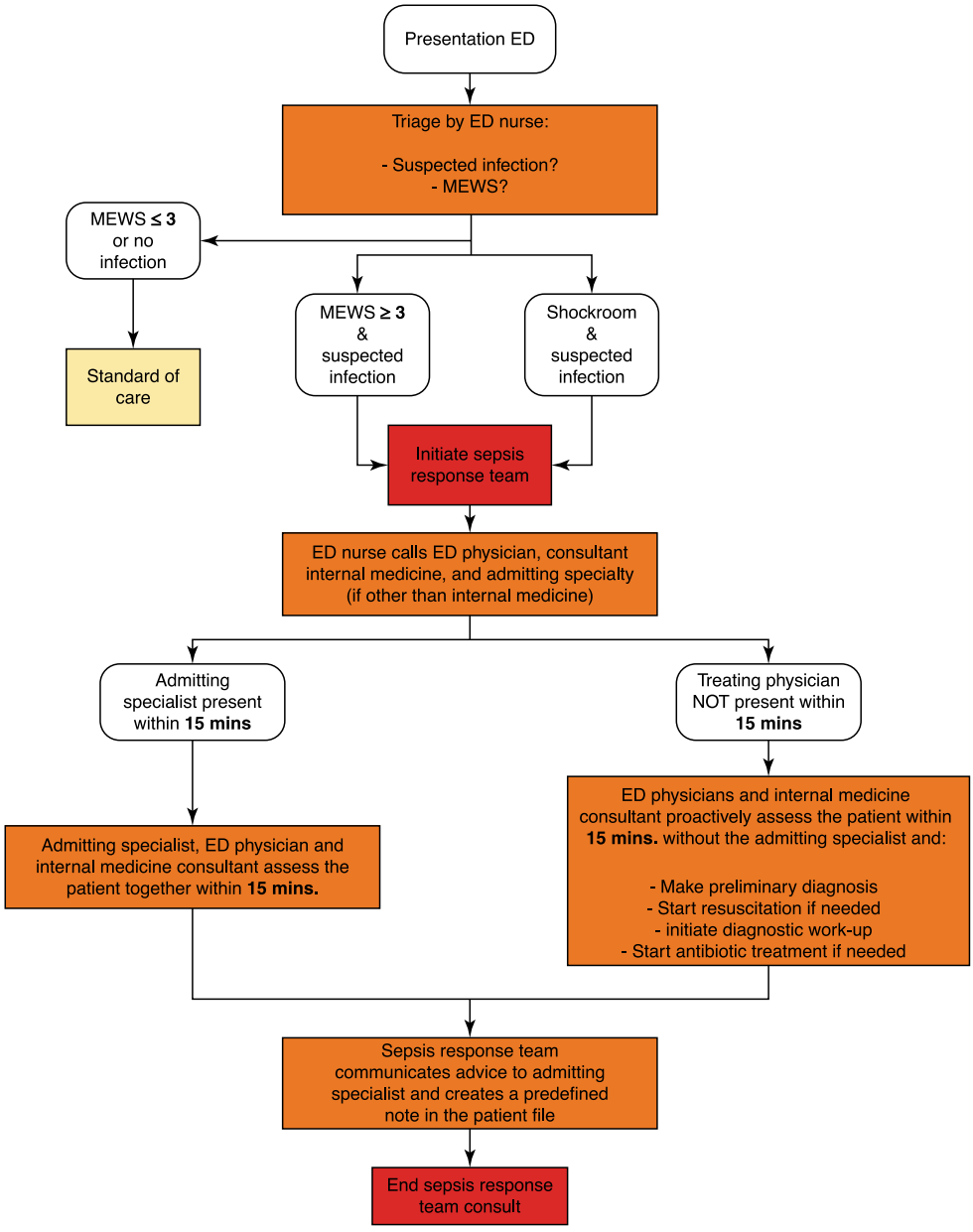



**Figure 1. Flowchart of sepsis response team involvement in a large teaching university medical center.** A practical example from Amsterdam University Medical Center including all aspects from early detection to the diagnostic work-up and treatment decisions. *ED* emergency department, *MEWS* Modified Early Warning Score

## CONCLUSION

Sepsis performance improvement programs can optimize compliance to sepsis care protocols, which have been associated with improved patient outcomes in various studies. These programs should ideally combine screening tools, process changes in sepsis care pathways, and educational programs to create awareness about sepsis care. The consequent gains through swift and adequate recognition of sepsis can be used to diagnose and treat patients accurately and timely according to (SSC) care protocols and deliberately think about when it is necessary to deviate from the general recommendations. Trust and behavior change are essential aspects of implementing sepsis care bundles. These aspects can be reinforced by performance improvement programs but need time. Engaging a large group of multidisciplinary clinical leaders for sepsis improvement programs seems essential for their success.

## ACKNOWLEDGEMENTS

This work is supported by an innovation grant from the Amsterdam UMC. We would also like to thank Idelette Nutma, patient representative and sepsis survivor, for contributing to this work.

## AUTHORS' CONTRIBUTIONS

MS and WJW conceived the study. MS, PWBN, and WJW analysed and interpreted the literature. MS drafted the work. PWBN and WJW substantially revised the work. MS, PWBN, and WJW all read and approved the final version of the manuscript and agree to be accountable for the integrity of the work. All authors read and approved the final manuscript.

## FUNDING

Publication costs were funded by the 2019 Amsterdam UMC Innovation Grant: "Het Sepsis Team: Betere Overleving Door Snelle Behandeling" (Project No: 23297).

## AVAILABILITY OF DATA AND MATERIAL

Not applicable.

## DECLARATIONS

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

PWBN and WJW declare that they have received an innovation grant from the Amsterdam UMC to study the effects of a sepsis improvement program. MS's Ph.D. studies are funded by this grant.

## Author details

[1]Center for Experimental and Molecular Medicine, Amsterdam UMC, Location Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [2]Section General Internal Medicine, Department of Internal Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC, Location VU University Medical Center, Amsterdam, The Netherlands. [3]Department of Medicine, Division of Infectious Diseases, Amsterdam UMC, Location Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES


1. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. 2021;47:1181–247.
2. Dellinger RP. The future of sepsis performance improvement. Crit Care Med. 2015;43:1787–9.
3. Viale P, Tedeschi S, Scudeller L, et al. Infectious diseases team for the early management of severe sepsis and septic shock in the emergency department. Clin Infect Dis. 2017;65:1253–9.
4. Djurkovic S, Baracaldo JC, Guerra JA, Sartorius J, Haupt MT. A survey of clinicians addressing the approach to the management of severe sepsis and septic shock in the United States. J Crit Care. 2010;25(658):e1-6.
5. Damiani E, Donati A, Serafini G, et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. PLoS ONE. 2015;10:e0125827.
6. Levy MM, Rhodes A, Evans LE, et al. COUNTERPOINT: should the surviving sepsis campaign guidelines be retired? No Chest. 2019;155:14–7.
7. Marik PE, Farkas JD, Spiegel R, et al. POINT: should the surviving sepsis campaign guidelines be retired? Yes Chest. 2019;155:12–4.
8. Gilbert DN, Kalil AC, Klompas M, Masur H, Winslow DL. IDSA position statement: why IDSA did not endorse the surviving sepsis campaign guidelines. Clin Infect Dis. 2017;45:486.
9. Spiegel R, Farkas JD, Rola P, Kenny JE, Olusanya S, Marik PE, Weingart SD. The 2018 surviving sepsis Campaign's treatment bundle: when guidelines outpace the evidence supporting their use. Ann Emerg Med. 2019;73:356–8.
10. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345:1368–77.
11. The ProCESS Investigators. A randomized trial of protocol-based care for early septic shock. N Engl J Med. 2014;370:1683–93.
12. The ARISE Investigators and the ANZICS Clinical Trials Group. Goal-directed resuscitation for patients with early septic shock. N Engl J Med. 2014;371:1496–506.
13. Angus DC, Barnato AE, Bell D, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. Intensive Care Med. 2015;41:1549–60.
14. Alam N, Oskam E, Stassen PM, et al. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. Lancet Respir Med. 2018;6:40–50.
15. Andrews B, Semler MW, Muchemwa L, et al. Effect of an early resuscitation protocol on in-hospital mortality among adults with sepsis and hypotension: a randomized clinical trial. JAMA. 2017;318:1233–40.
16. Rothrock SG, Cassidy DD, Barneck M, et al. Outcome of immediate versus early antibiotics in severe sepsis and septic shock: a systematic review and meta-analysis. Ann Emerg Med. 2020;76:427–41.
17. Schinkel M, Nannan Panday RS, Wiersinga WJ, Nanayakkara PWB. Timeliness of antibiotics for patients with sepsis and septic shock. J Thorac Dis. 2020;12(Suppl 1):S66-71.
18. Levy MM, Rhodes A, Phillips GS, et al. Surviving sepsis campaign: association between performance metrics and outcomes in a 7.5-year study. Crit Care Med. 2015;43:3–12.
19. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). Crit Care. 2010;14:R8.
20. Miller RR, Dong L, Nelson NC, et al. Multicenter implementation of a severe sepsis and septic shock treatment bundle. Am J Respir Crit Care Med. 2013;188:77–82.
21. Castellanos-Ortega Á, Suberviola B, García-Astudillo LA, Holanda MS, Ortiz F, Llorca J, Delgado-Rodríguez M. Impact of the surviving sepsis campaign protocols on hospital length of stay and mortality in septic shock patients: results

of a three-year follow-up quasi-experimental study. Crit Care Med. 2010;38:1036–43.

22. Van Zanten ARH, Brinkman S, Arbous MS, Abu-Hanna A, Levy MM, De Keizer NF. Guideline bundles adherence and mortality in severe sepsis and septic shock. Crit Care Med. 2014;42:1890–8.
23. Varpula M, Karlsson S, Parviainen I, Ruokonen E, Pettilä V. Community-acquired septic shock: early management and outcome in a nationwide study in Finland. Acta Anaesthesiol Scand. 2007;51:1320–6.
24. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, Shieh L. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Qual Saf. 2019;28:762–8.
25. Hooper MH, Weavind L, Wheeler AP, et al. Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit. Crit Care Med. 2012;40:2096–101.
26. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4:e000234.
27. Nannan Panday RS, Minderhoud TC, Alam N, Nanayakkara PWB. Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): a narrative review. Eur J Intern Med. 2017;45:20–31.
28. Liu VX, Lu Y, Carey KA, et al. Comparison of early warning scoring systems for hospitalized patients with and without infection at risk for in-hospital mortality and transfer to the intensive care unit. JAMA Netw Open. 2020;3:e205191.
29. Goulden R, Hoyle MC, Monis J, et al. QSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. Emerg Med J. 2018;35:345–9.
30. Yu SC, Shivakumar N, Betthauser K, et al. Comparison of early warning scores for sepsis early identification and prediction in the general ward setting. JAMIA Open. 2021;4:1–6.
31. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the emergency department. Am J Emerg Med. 2019;37:1490–7.
32. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Prog Biomed. 2019;170:1–9.
33. Van Wyk F, Khojandi A, Kamaleswaran R. Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study. IEEE J Biomed Heal Inform. 2019;23:978–86.
34. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: a narrative review. Comput Biol Med. 2019;115:103488.
35. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181:1065–70.
36. Girardis M, Rinaldi L, Donno L, Marietta M, Codeluppi M, Marchegiano P. Venturelli C (2009) effects on management and outcome of severe sepsis and septic shock patients admitted to the intensive care unit after implementation of a sepsis program: a pilot study. Crit Care. 2009;135:1–8.
37. Tromp M, Hulscher M, Bleeker-Rovers CP, et al. The role of nurses in the recognition and treatment of patients with sepsis in the emergency department: a prospective before-and-after intervention study. Int J Nurs Stud. 2010;47:1464–73.
38. Nguyen HM, Schiavoni A, Scott KD, Tanios MA. Implementation of sepsis management guideline in a community-based teaching hospital—can education be potentially beneficial for septic patients? Int J Clin Pract. 2012;66:705–10.
39. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315:801.
40. Marshall JG, Vincent JL, Guyatt G, et al. Outcome measures for clinical research

in sepsis: a report of the 2nd Cambridge colloquium of the international sepsis forum. Crit Care Med. 2005;33:1708–16.

41. Marshall JC, Murthy S, Diaz J, et al. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis. 2020;20:e192–7.

42. Schinkel M, Virk HS, Nanayakkara PWB, van der Poll T, Wiersinga WJ. What sepsis researchers can learn from COVID-19. Am J Respir Crit Care Med. 2021;203:125–7.

43. Hartzband P, Groopman J. There is more to life than death. N Engl J Med. 2012;367:987–9.

# CHAPTER 3

## THE IMPACT OF A SEPSIS PERFORMANCE IMPROVEMENT PROGRAM IN THE EMERGENCY DEPARTMENT: A BEFORE-AFTER INTERVENTION STUDY

Michiel Schinkel[1,2,3], Frits Holleman[4], Richarda Vleghels[1,6], Kayla Brugman[2], Milan L. Ridderikhof[5], Mahi Dzelili[6], Prabath W.B. Nanayakkara[2,7] & W. Joost Wiersinga[1,3,4]

1. Amsterdam UMC location University of Amsterdam, Center for Experimental and Molecular Medicine (CEMM), Meibergdreef 9, Amsterdam, the Netherlands
2. Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Internal Medicine, De Boelelaan 1117, Amsterdam, The Netherlands.
3. Amsterdam Institute for Infection and Immunity (AII), Amsterdam, the Netherlands.
4. Amsterdam UMC University of Amsterdam, Department of Internal Medicine, Meibergdreef 9, Amsterdam, The Netherlands.
5. Amsterdam UMC University of Amsterdam, Emergency Department, Meibergdreef 9, Amsterdam, The Netherlands.
6. Amsterdam UMC location Vrije Universiteit Amsterdam, Emergency Department, De Boelelaan 1117, Amsterdam, The Netherlands.
7. Amsterdam Public Health, Amsterdam, the Netherlands

*Infection*
*https://doi.org/10.1007/s15010-022-01957-x*

# ABSTRACT

## Purpose

The latest Surviving Sepsis Campaign guidelines advocate that all hospitals use sepsis performance improvement programs. However, there is limited evidence about how to structure such programs and what their potential impact is on sepsis management and outcomes in the emergency department (ED). In this study, we evaluated the implementation of a sepsis performance improvement program in the ED including a dedicated sepsis response team and analyzed the management and outcomes of sepsis patients before and after.

## Methods

We conducted a before-after interventional study in the ED of the Amsterdam University Medical Centers, the Netherlands. The sepsis performance improvement program included regular educational meetings, daily audits and weekly feedback, a screening tool, and a dedicated multidisciplinary sepsis response team. We studied all adult patients who presented to the ED with a suspected infection and a Modified Early Warning Score (MEWS) ³ 3 during their stay. In the postintervention phase, these patients were seen by the sepsis team. Process-related and patient-related outcomes were measured between November 2019 – February 2020 (preintervention) and December 2021- May 2022 (postintervention).

## Results

A total of 265 patients were included in the primary study, 132 patients preintervention and 133 patients postintervention. The postintervention phase was associated with improvements in nearly all process-related outcomes, such as a shorter time to antibiotics (66 vs. 143 minutes; p<0.001), increased number of lactate measurements (72.9% vs. 46.2%; p<0.001), and improved completeness of documented MEWS scores (85.0% vs. 62.9%; p<0.001). Except for an improvement in the number of immediate versus delayed ICU admissions (100% immediate vs. 64.3% immediate; p=0.012), there was no improvement in the other patient-related outcomes such as 28-day mortality (14.3% vs. 9.1%; p=0.261), during the postintervention phase.

## Conclusion

Our program stimulated physicians to make timely decisions regarding diagnostics and treatment of sepsis in the ED. Implementing the sepsis performance improvement program was associated with significant improvements in most process-related outcomes but with minimal improvements in patient-related outcomes in our cohort.



## Keywords

Sepsis; Surviving Sepsis Campaign; Sepsis Performance Improvement Program; Sepsis Team

## INTRODUCTION

Sepsis is a major global health problem defined as a life-threatening organ dysfunction caused by a dysregulated host response to infection [1, 2]. In 2017, the World Health Assembly of the World Health Organization declared sepsis a global priority and adopted a resolution to improve its prevention, diagnosis, and management [3]. With a recent estimate of 49 million sepsis cases each year, the global burden of sepsis may be more significant than previously anticipated [4, 5].

In the early 2000s, the Surviving Sepsis Campaign (SSC) was established to provide evidence-based guidelines for managing sepsis and septic shock [6]. The SSCs goal is to reduce sepsis morbidity and mortality worldwide, and its guidelines were most recently updated in 2021 [7, 8]. By bundling the guideline recommendations into core groups of clinical actions that should be performed within a specific timeframe, the SSC aims to facilitate implementation [8]. Several observational studies have shown that compliance with sepsis care bundles is associated with reduced mortality rates [9–13]. However, bundle adherence still remains a significant challenge, and non-compliance is especially prominent in the microbiological workup and timely administration of antibiotics in emergency departments (EDs) [14–16]. As the ED often represents a sepsis patient's first interaction with the healthcare system, it is crucial to promptly initiate the appropriate care processes in this setting [16].

In response to the suboptimal compliance with sepsis guidelines, hospitals and healthcare organizations have initiated sepsis performance improvement programs. These initiatives often include interventions such as educational programs, screening tools, or changes in sepsis care pathways (e.g., activating dedicated sepsis response teams) [17]. Performance improvement programs have been associated with better adherence to SSC or local sepsis guidelines and decreased mortality rates [18]. The latest SSC guideline thus advocates that all hospitals and health systems implement sepsis performance improvement programs [8].

Although the use of sepsis performance improvement programs is now recommended, there is limited evidence on how these programs should optimally be structured [17, 18]. Furthermore, their potential impact on the ED population is relatively unknown, as most studies target the intensive care unit (ICU) population [17, 18]. In this study, we prospectively evaluate the implementation of a multidisciplinary sepsis response team and performance improvement program in our ED and analyze the management and outcomes of sepsis patients before and after.

## METHODS

We conducted a before-after intervention study in the ED of the Amsterdam University Medical Centers - location VUmc, in the Netherlands. The Medical Ethics Review Committee waived the review of this study as it was a quality improvement project within regular care (IRB number: IRB00002991; case:19.449). Study outcomes were measured between November 2019 – February 2020 (preintervention phase) and December 2021 - May 2022 (postintervention phase), while the period in between (March 2020 – November 2021) was used to implement all aspects of the sepsis performance improvement program appropriately (implementation phase).Patients were sent a letter to opt out of the use of their data for this project. We adhere to "The Strengthening

the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies" [19].

## Study population



We studied all adult patients (18 years and older) who presented to the ED with a suspected infection and a Modified Early Warning Score (MEWS) ≥ 3 during their stay. We used the MEWS with a cut-off at three to screen for sepsis, following the guidelines from the Dutch Federation of Medical Specialists (FMS), National Patient Safety Programme (VMS), and the SSC guideline [8, 20, 21]. Patients were excluded from the study if they were pregnant, SARS-CoV-2 positive before arriving at the ED, or when they opted out of participating.

## Intervention

The preintervention measurements were performed before our sepsis performance improvement initiative was started, and the hospital's standard care was provided to all patients with suspected infections. The MEWS was already part of the hospital's standard screening procedures during this period, although compliance with these procedures was variable. Afterwards, we introduced several interventions: regular educational meetings at morning handovers; standardized sepsis team notes and order sets in the Electronic Health Record (EHR) system; daily audits and weekly feedback; the systematic use of a screening tool (MEWS); and most importantly, the introduction of a multidisciplinary sepsis response team. During the postintervention phase, the sepsis team was active in the ED 24-hours a day, seven days a week. An ED nurse on duty alerted the sepsis team when a patient was identified as having a suspected infection and had a MEWS ≥ 3 during the ED stay. The multidisciplinary sepsis response team consisted of the on-call physician from the following departments: emergency medicine, internal medicine, and the admitting specialty (e.g., surgery, urology, neurology, etc.). The team aimed to assess all patients within 15 minutes after a MEWS ≥ 3 was recorded in the ED. Following the assessment of the patient, the team advised the on-call physician of the admitting specialty regarding the diagnostic workup and treatment based on the local protocol, which was adapted according to the SSC guidelines. The Amsterdam UMC follows the national antibiotic sepsis guidelines of the Dutch Working Party on Antibiotic Policy (SWAB; https://swab.nl/en/swab-guidelines), which did not change during the study period. The workflow, which focused on collaboration and shared responsibility across specialties, was created with input from emergency and intensive care physicians, internal medicine specialists, radiologists, and patient representatives. The complete sepsis team workflow is visually presented in Figure 1.

## Data and outcomes

To study the impact of the implementation of our sepsis team, we looked at two distinct data categories: process-related and patient-related outcomes. All study data were collected from the EHR. The base dataset included patient characteristics such as age, sex, and comorbidities. For process-related outcomes, we collected data on aspects of the diagnostic workup (e.g., blood cultures taken, lactate measurements) and the treatment strategy (e.g., administration of antibiotics

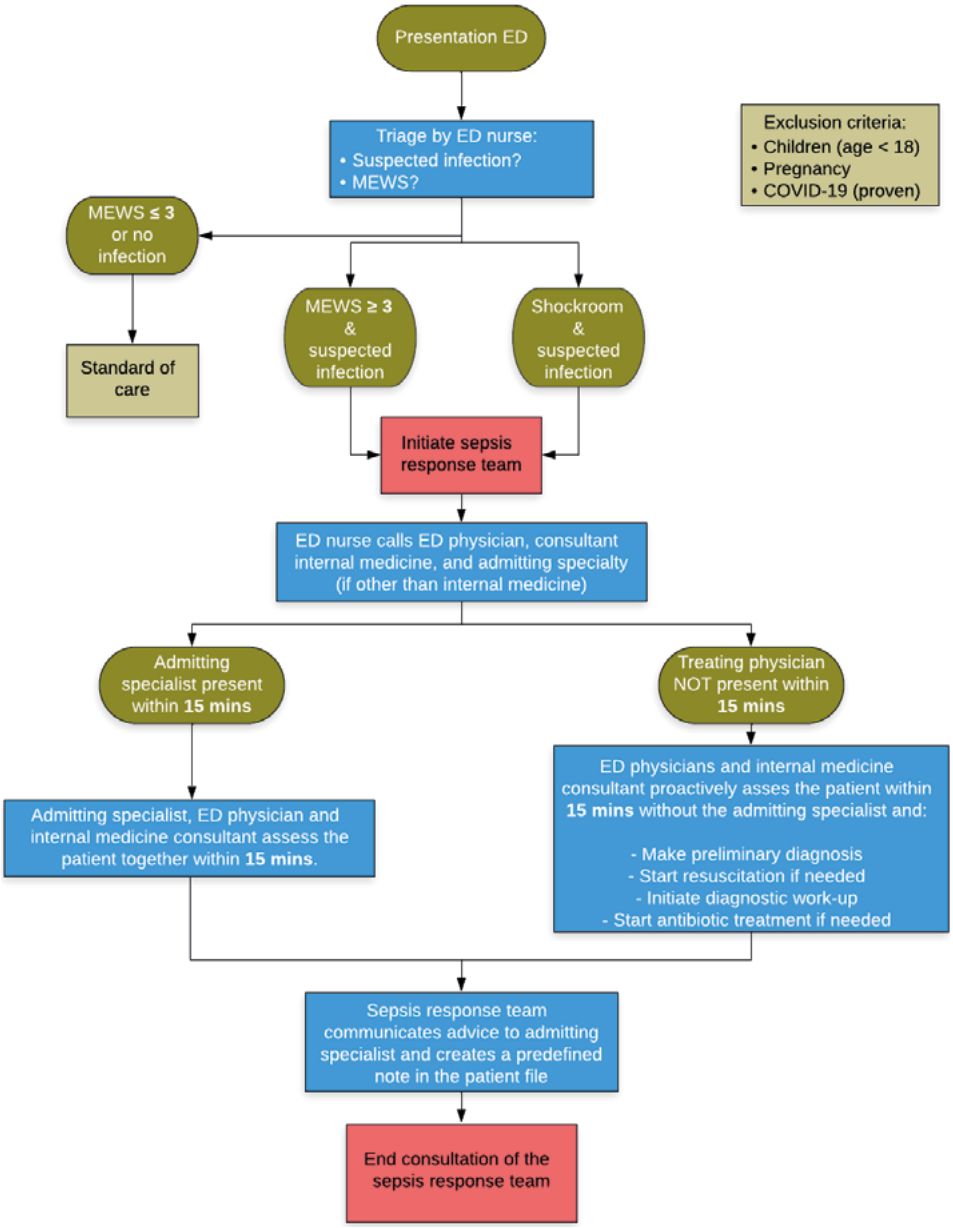



**Figure 1. The flowchart for the activation of the sepsis response team as implemented in the ED of the Amsterdam UMC.** ED = Emergency Department; MEWS = Modified Early Warning Score.

and fluids). For patient-related outcomes, we extracted data such as ED length of stay, admission rates, and mortality rates. Furthermore, we studied the number of patients directly admitted to the ICU from the ED (immediate ICU admission) compared with those who were first admitted to regular wards and further transferred to the ICU within the first 48 hours (delayed ICU admission).

Besides the outcome measures, we also investigated the specific infections and diagnoses in the study cohort. The type of infection and final diagnosis were assessed by a clinical adjudication committee consisting of an experienced ED nurse (RV), a last-year medical student (KB), and a medical doctor (MS), and based on all microbiology results (including all culture results, polymerase chain reaction testing, etc) and the medical notes in the EHR. Furthermore, sepsis or the progression to sepsis in the first 72 hours was assessed based on the Sepsis-3 definition, using the Sequential Organ Failure Assessment (SOFA) score [1]. The SOFA score uses PaO2/FiO2 ratios to determine respiratory dysfunction. These ratios require arterial blood gas measurements, which are infrequently performed in EDs. Therefore, we used the SpO2/FiO2 ratio and corresponding cut-offs as a proxy for the PaO2/FiO2 ratio, as previously described [22, 23]. For all SOFA items, the worst value of the day was used. When an item was not measured on a given day, the score for that part of the SOFA was zero.

### Statistical analysis

We hypothesized that the ED length of stay would be the most likely patient-related outcome that could be impacted by introducing a sepsis team. To detect a statistically significant change in the ED length of stay of at least 30 minutes, from a retrospective baseline of approximately 228 minutes and a standard deviation of 87 minutes, we needed to include a total of 266 patients (133 patients per phase) to find a difference with a power of 80%.

Descriptive statistics were used to assess all variables. Continuous variables were described by their median and interquartile range (IQR) and compared between the preintervention and postintervention phase using a T-test or Wilcoxon-Mann-Whitney test when appropriate considering the distribution. Categorical variables were described using frequencies and percentages, and differences were calculated using the Chi-square test.

During the postintervention phase, the sepsis team did not cover all eligible patients suspected of infection with an elevated MEWS score. Therefore, a comparison between the postintervention groups "sepsis team activated" and "sepsis team not activated" was conducted to better evaluate the sepsis team's effect. Group differences were further examined with a multivariable linear or logistic regression to establish whether these differences could be explained by differences in baseline characteristics between the groups. The outcomes were adjusted for age, comorbidity index, MEWS in the ED, and "do not resuscitate" policies.

Statistical significance was defined as a two-tailed p-value < 0.05. The analyses were performed using R (version 4.2.1) and the R packages: "tidyverse", "ggplot2", "ggpubr", "naniar", and "tableone" [24].

## RESULTS

A total of 265 patients were included in the primary study, 132 patients preintervention and 133 postintervention. The median age was 68 years (interquartile range [IQR]: 56 - 77), and 61.5% of the patients were male. The median MEWS score in the ED was 4 (IQR: 4-6), and the Charlson Comorbidity Index was 5 (IQR: 2-7). Baseline characteristics are provided in Table 1. The most common site of the suspected Infection in the ED was the respiratory tract in both the pre and

**Table 1.** Characteristics of patients presenting to the emergency department with a suspected infection and Modified Early Warning Score ≥3. A comparison is made between the preintervention and postintervention phases of the implementation of the sepsis improvement program in the ED.

| Characteristics | Totals (n=265) | Preintervention (n=132) | Postintervention (n=133) | P-value |
|---|---|---|---|---|
| Age, year | 68 (56-77) | 69 (57.5-77) | 67 (56-76) | 0.574 |
| Sex, male | 163 (61.5%) | 80 (60.6%) | 83 (62.4%) | 0.861 |
| Charlson Comorbidity Index | 5 (2-7) | 5 (2-7) | 5 (3-7) | 0.367 |
| MEWS in ED | 4 (3-6) | 4 (3-5) | 5 (4-6) | <0.001 |
| Do not resuscitate orders | 89 (33.7%) | 38 (29.0%) | 51 (38.3%) | 0.275 |
| **Diagnostic workup and treatment** | | | | |
| Complete MEWS recorded | 196 (74.0%) | 83 (62.9%) | 113 (85.0%) | <0.001 |
| Time to MEWS ≥ 3, minutes | 17 (9-30) | 19 (10-28.75) | 16 (8-30) | 0.315 |
| Blood culture taken in the ED | 209 (78.9%) | 93 (70.5%) | 116 (87.2%) | <0.001 |
| Time to blood culture, minutes | 28 (14-65) | 42 (20-126) | 24 (9.75-49.75) | <0.001 |
| Antibiotics administered in the ED | 179 (67.5%) | 73 (55.3%) | 106 (79.7%) | <0.001 |
| Time to first antibiotics, minutes | 95 (43-181) | 143 (91-250) | 66 (40-123.75) | <0.001 |
| Antibiotics administered and blood culture taken in the ED | 159 (60.0%) | 62 (47.0%) | 97 (72.9%) | <0.001 |
| Blood cultures before antibiotics | 138 (86.8%) | 52 (83.9%) | 86 (88.7%) | <0.001 |
| Lactate measurement | 158 (59.6%) | 61 (46.2%) | 97 (72.9%) | <0.001 |
| Repeat measurement of lactate | 52 (32.9%) | 10 (16.4%) | 42 (43.3%) | <0.001 |

ED = Emergency Department; MEWS = Modified Early Warning Score. Data are presented with no. (%) or median (interquartile range). The results of the indented characteristics are calculated based on the number of patients in the main group of that characteristic.

post-intervention phases. However, respiratory tract infections were relatively more common in the preintervention phase. e-Figure 1 of the supplementary appendix shows the distribution of the suspected infection sites and the most likely infection type at discharge in the different phases.

## Before-after comparison

### *Process-related*

First, we studied the effect of our sepsis performance improvement program on process-related outcomes. During the postintervention phase, the complete MEWS assessment (all items recorded) was performed more frequently (85.0% vs. 62.9%; p<0.001), but the time to the first recorded MEWS (≥ 3) was similar (16 vs. 19 min; p=0.315). Blood cultures were drawn significantly more often during the postintervention phase (87.2% vs. 70.5%; p<0.001), and the time until blood cultures were drawn was lower (24 vs. 42 minutes; p<0.001). Antibiotics were administrated in the ED in more cases during the postintervention phase (79.7% vs. 55.3%; p<0.001), and the time to antibiotics was significantly lower (66 vs. 143 minutes; p<0.001). Lactate measurements were performed more often (72.9% vs. 46.2%; p<0.001), and repeat measurements were also performed more frequently (43.3% vs. 16.4%; p<0.001). Taken together, these results show that the implementation of our sepsis performance improvement program was associated with improvements in most process indicators.

### *Patient-related outcomes*

Next, we examined whether the process-related improvements translated into improved patient-related outcomes. Our main outcome parameter was the length of stay in the ED, which was similar in the postintervention phase compared to the preintervention phase (283 vs. 287 minutes; p=0.983). Hospital admission rates were also similar in both phases (88.0% vs. 83.3%; p=0.367). There were no significant differences in the number of ICU admissions (16.5% vs. 10.6%; p=0.353) or the length of stay in the ICU (2 vs. 2.5 days; p=0.830). However, the number of immediate ICU admissions from the ED was significantly higher in the postintervention group compared to the preintervention group (100% immediate vs. 64.3% immediate; p=0.012). However, the hospital length of stay was significantly longer in the postintervention phase (5 days vs. 4 days; p=0.033). There were no significant differences regarding 28-day mortality (14.3% vs. 9.1%; p=0.261) or 28-day hospital readmissions (10.5% vs. 17.4%; p=0.096) between the groups. An overview of the patient-related outcomes is provided in Table 2.

To further investigate whether the implementation of our sepsis team may have impacted mortality, we created a logistic regression model to explain 28-day mortality by the sepsis team implementation, adjusted for age, MEWS, comorbidity index, and "do not resuscitate" (DNR) policy. The odds ratio (OR) for 28-day mortality in the postintervention phase was 1.24 (95% confidence interval (CI): 0.54-2.92; p=0.611). In this model, only the DNR policy was significantly associated with 28-day mortality with an OR of 6.70 (95% CI: 2.53-20.12; p<0.001). We also created a linear regression model to examine whether the significantly longer length of stay in the hospital in patients seen by the sepsis team could be explained by differences in the baseline characteristics. After adjustment for age, MEWS, comorbidity index, and DNR policy, the use of the sepsis team was no longer associated with a prolonged hospital stay (p=0.171). In this model, only the MEWS in the ED was significantly associated with the hospital length of stay (p<0.001). Overall, these results show that,

**Table 2.** Outcomes of patients presenting to the emergency department with a suspected infection and Modified Early Warning Score ³3. A comparison is made between the preintervention and postintervention phases of the implementation of the sepsis improvement program in the ED.

| Outcome | Totals (n=265) | Preintervention (n=132) | Postintervention (n=133) | P-value |
|---|---|---|---|---|
| ED length of stay, minutes | 286 (221-407) | 287 (224-407) | 283 (221-409) | 0.983 |
| Hospital admission | 227 (85.7%) | 110 (83.3%) | 117 (88.0%) | 0.367 |
| Hospital length of stay, days | 5 (3-9) | 4 (2-9) | 5 (3-10) | 0.033 |
| ICU admission | 36 (13.6%) | 14 (10.6%) | 22 (16.5%) | 0.456 |
| Immediate ICU admission from ED | 31 (86.1%) | 9 (64.3%) | 22 (100%) | 0.012 |
| ICU length of stay, days | 2 (1.75-7.75) | 2.5 (2-4) | 2 (1.25-11.50) | 0.704 |
| 28-day mortality | 31 (11.7%) | 12 (9.1%) | 19 (14.3%) | 0.261 |
| 28-day readmission | 37 (14.0%) | 23 (17.4%) | 14 (10.5%) | 0.096 |
| SOFA score ≥ 2 within 72 hours | 193 (72.8%) | 95 (72.0%) | 98 (73.7%) | 0.861 |

ED = Emergency Department; ICU = Intensive Care Unit; SOFA = Sequential Organ Failure Assessment. Data are presented with no. (%) or median (interquartile range). The results of the indented characteristics are calculated based on the number of patients in the main group of that characteristic.

except for an increase in the percentage of direct ICU admissions from the ED compared to delayed ICU admissions, there were no meaningful differences in patient-related outcomes before and after the implementation of our sepsis performance improvement program.

### *Infection and sepsis*



We also investigated the type of patients identified through our intervention program. The number of patients who fulfilled the sepsis criteria during the first 72 hours of admission was calculated using the SOFA score. We found no differences in the number of patients fulfilling the Sepsis-3 criteria (73.7% vs. 72.0%; p=0.395) in the preintervention and postintervention phases. When we looked at the most likely etiology of the infections at discharge (based on all microbiology results and medical notes), we observed different distributions of causative agents before and after the implementation of the sepsis team. As shown in e-Figure 1, the most common preimplementation infection type was viral (non-COVID-19; predominantly influenza). After the implementation, the majority of infections were bacterial.

## Comparison between the postintervention groups (post-post)

Since there was no complete compliance with the sepsis team activation, we could study an additional cohort of control patients in the postintervention phase for whom the sepsis team was not activated. During the postintervention phase, the sepsis team was activated for 133/207 (64%) of all eligible patients. A comparison of the baseline characteristics of postintervention groups in which the sepsis team was, or was not activated, is shown in e-Table 1 of the supplementary appendix.

The patients for whom the sepsis team was activated had similar Charlson Comorbidity Index scores (5 vs. 4; p=0.078) but a higher MEWS score on presentation (5 vs. 4; p=0.007). When the sepsis team was activated, MEWS scores were recorded completely in more cases (85.0% vs. 66.2%; p=0.003). A similar number of blood cultures was performed (87.2% vs. 86.5%; p=1.000), but they were performed faster when the sepsis team was activated (24 vs. 43.5 minutes; p=0.009). Antibiotics were administered more frequently (79.7% vs. 63.5%; p=0.017), and the time to antibiotic treatment was lower (66 vs. 126 minutes; p=0.001) in those patients for which the sepsis team was activated compared to those patients for which the sepsis team was not activated. Activation of the sepsis team resulted in a higher number of lactate measurements (72.9% vs. 50.0%; p=0.002), while the rates of repeat measurement lactate levels were statistically comparable (43.3% vs. 24.3%; p=0.085).

Except for an increased number of ICU admissions directly from the ED (p=0.033), we observed no significant differences in patient-related outcomes such as ED length of stay, admission rates, or mortality rates, as further highlighted in e-Table 2 of the supplementary appendix. However, the postintervention group in which the sepsis team was activated had significantly more cases that fulfilled the sepsis criteria within the first three days of admission (73.7% vs. 51.4%; p=0.002). Taken together, the post-post comparison reinforces that the sepsis improvement program is associated with improved process-related outcomes, including a 50% lower time to antibiotics, though there were few observable patient benefits.

## DISCUSSION

In this study, we found that the implementation of a sepsis performance improvement program in the ED including the use of a specialized multidisciplinary sepsis response team resulted in better identification of sepsis, an improved diagnostic process, and a >50% reduction in time to antibiotic treatment in suspected sepsis patients. However, these improved process-related outcomes did not translate into improvements in length of stay, admission rates, or mortality rates. The only patient-related outcome which improved was the number of immediate versus delayed ICU admissions.

Implementing our sepsis performance improvement program was associated with various improved process-related outcomes. The MEWS score, which already was the preferred screening tool according to hospital policy in the preimplementation phase, was recorded completely in significantly more cases. Interestingly, the MEWS recordings were also more complete when comparing postimplementation patients for whom the sepsis team was or was not activated (post-post comparison). This indicates that the triage nurses indeed linked the use of the MEWS as a screening tool to the sepsis performance improvement program but did not use it when they did not consider the patient as a potential sepsis case. In addition, the number of lactate measurements also significantly increased after implementation and remained significant in the post-post comparison. The SSC guideline recommends using both sepsis screening tools (including MEWS) and lactate measurements, and we thus show increased SSC guideline adherence in these instances [8]. Regarding the workup with blood cultures, we found mixed results. Though there seems to be an increase in the number of blood cultures performed when comparing preimplementation and postimplementation patients, the post-post comparison does not reinforce this effect. ED nurses seem to have been more inclined to draw blood cultures postintervention, irrespective of whether or not the patient was seen by the sepsis team. This could be due to the attention given to blood cultures in the educational meetings as part of the sepsis improvement program, but it could also be a reflection of the higher rate of bacterial infections in the post-phase, as seen in Figure 1B of the supplementary appendix. In line, sepsis team utilization was associated with a decrease in time to blood culture draws when compared to both control groups.

The implementation of the sepsis team was associated with a considerable reduction in the time to the first administration of antibiotics. Although the benefits of early antibiotics for all sepsis patients remain debatable, there are specific subgroups of patients who may experience benefits [25–29]. The sepsis team also started antibiotic treatment in the ED in significantly more cases than in the preimplementation phase or in the postimplementation patients when no sepsis team was involved. This indicates that the improved recognition of sepsis may have led to an increased use of antibiotics. Notably, not all patients seen by the sepsis team were treated with antibiotics in the ED. In cases with a relatively low probability of sepsis or shock, the latest SSC guidelines suggest conducting a time-limited investigation first and only initiating antimicrobial therapy when the concern for infection persists [8].

Despite many improved process-related outcomes, we found only a single patient-related outcome that improved after implementing the sepsis team in the ED, which contrasts with previous literature [16, 30, 31]. Viale *et al.* found that their infectious diseases team improved SSC

guideline adherence in a general ED in Italy [16]. Their pre-post comparison including 382 (195 vs. 187) severe sepsis and septic shock patients with a high median age of 82 years (IQR 70-88) showed that the infectious diseases team implementation was associated with higher rates of lactate measurements (90% vs. 76%; p<0.001) and blood cultures before antibiotics (58% vs. 42%; p<0.001). The time to first antibiotic treatment did not significantly decrease (154 vs. 169 minutes; p=0.42). Interestingly, the all-cause 14-day mortality was significantly lower in univariate and multivariate analyses (29% vs 39%; p=0.02), but the 30-day all-cause mortality was not (37% vs. 45%; p=0.102). Arabi *et al*. implemented a multifaceted intervention similar to ours, including a sepsis response team, in their ED in Saudi Arabia [31]. In that postintervention cohort of 699 patients, most process-related and patient-related outcomes improved significantly. For example, the percentage of patients receiving antibiotics within three hours improved (89.4% vs. 67.7%; p<0.001), and the hospital mortality rate was lower (16.9% vs. 47.7%; p=0.003). A recent publication by Simon *et al*. also shows improvements associated with a sepsis team implementation in the ED of a tertiary hospital in the United States. The pre-post analysis among 863 patients (393 vs. 470) showed that the time to antibiotics was reduced (81 vs. 107 minutes; p<0.001), just as the in-hospital mortality (15.1% vs. 28.2%; p<0.001). A notable difference with all of these cohorts is that their preintervention mortality rates of 45%, 48%, and 28% were much higher than in our cohort (9.1%). This finding is not completely unexpected since other Dutch studies and various international sepsis studies in the ED setting have also reported relatively low sepsis mortality rates [26, 27]. Furthermore, the aim of this sepsis performance improvement program was to screen for sepsis and detect and treat it early. Mortality rates in such a screening cohort will be lower than in cohorts looking only at definite and severe sepsis cases. We may argue that sepsis performance improvement programs are more likely to improve mortality at those higher mortality rates or that establishing a significant mortality benefit is at least easier in such a population. Still, we would have expected to find other improvements in patient-related outcomes through our intervention, especially a shorter length of ED stay. Unfortunately, overcrowding of the ED and exit blocks toward the hospital wards due to staff shortages in our postintervention phase made it challenging to transfer patients to the wards [32, 33]. During the extraction of data from the EHR system, the study team had the impression that patients were ready for hospital admission earlier when they were seen by the sepsis team, but this could not be reflected in shorter ED stays due to the logistical constraints. This hypothesis is further supported by the fact that we were able to show a significant improvement in the number of direct versus delayed ICU admissions. Delayed ICU admission (patients who will eventually need an ICU admission but are first admitted to the regular ward) is an independent risk factor for sepsis mortality, but none of the patients seen by the sepsis team were being admitted to the ICU with a delay [34, 35]. This suggests that our intervention helped bring together the experts needed to make the most appropriate and timely decision about where the patient needed to go next.

To fully understand the results of this study, it is essential to acknowledge the role of the COVID-19 pandemic. Shortly after our preimplementation measurements, the SARS-CoV-2 virus emerged [36]. The pandemic put unprecedented pressure on healthcare workers and hospitals, which caused a significant delay in the implementation of our sepsis team [37]. Consequently, a near two-year interval was needed before the postimplementation measurements could be performed.

Even then, the healthcare system, and certainly the ED, continued to operate under high pressure, which led to imperfect compliance rates with the sepsis team activation. In the meantime, the national report on infectious diseases in the Netherlands and several international publications showed that the distribution of infectious agents had changed, with, for example, a much lower prevalence of influenza [38–40]. Our study observed similar changes, where influenza was much less prevalent in the postimplementation phase, while bacterial infection rates were higher. Of note, proven COVID-19 cases at presentation were excluded. Although sepsis guidelines are created for a heterogeneous group of patients with all types of infections, the results of our before-after comparison may have been influenced by the COVID-19 pandemic-induced changes in the causative agents of sepsis in our population. Bacterial infections seemed to have been much more prevalent during the postintervention phase. Fortunately, the imperfect compliance rates with the sepsis team activation led us to have an additional cohort of patients from the postimplementation phase in whom no sepsis team was activated and who could serve as an unexpected but essential second control group.

Besides the potential confounding through COVID-19, several other limitations of this study must be addressed. First and foremost, this study was powered to detect a difference in the length of ED stay after implementing a sepsis team. In hindsight, this was an unattainable result due to the logistical constraints (e.g., exit blocks) discussed above. Secondly, the compliance rate with the sepsis team activation was only 64%. Consequently, selection bias may have been introduced since the ED nurses may have had an unconscious bias to activate the sepsis team only in more severe cases, in whom the diagnostic workup and start of treatment would already happen more timely. Fortunately, we could negate part of this confounding by comparing the pre-post results to the post-post comparison. Still, the fact that patients in the postintervention phase may have been more severely ill and less likely to survive compared with the preintervention phase should be considered when interpreting these results. Interestingly, it seems that the ED nurses could identify the patients with a higher likelihood of having sepsis, as the rate of progression to sepsis was significantly higher in postimplementation patients who were seen by the sepsis team. This finding supports our approach of implementing a sepsis response workflow based on the SSC recommendations but with relative flexibility to maneuver according to clinical judgment. Lastly, the before-after study design has its inherent limitations, such as time-related changes in populations and standards of care. A large (stepped wedge cluster) randomized trial is needed to fully understand the value of sepsis teams and sepsis performance improvement programs in general. Given the limited evidence for the benefits and the proper structure of a sepsis team, we did not have the support base to conduct such a trial. We hope our current work helps create the urgency for this type of study.

In conclusion, implementing our sepsis performance improvement program in the ED was associated with a number of improvements in process-related outcomes but minimal improvements in patient outcomes. The program stimulated physicians to make collaborative and timely decisions regarding diagnostics and treatment of sepsis. The workflow allowed them to incorporate their clinical judgment while still reinforcing the essential elements of sepsis care.

## LIST OF ABBREVIATIONS

ED = Emergency Department
EHR = Electronic Health Records
FiO2 = Fraction of Inspired Oxygen
ICU = Intensive Care Unit
IQR = Interquartile Range
IRB = Institutional Review Board
MEWS = Modified Early Warning Score
PaO2 = Partial Pressure of Oxygen
PCR = Polymerase Chain Reaction
SOFA = Sequential Organ Failure Assessment
SpO2 = Peripheral Oxygen Saturation
SSC = Surviving Sepsis Campaign
STROBE = Strengthening of the Reporting of Observational Studies in Epidemiology

## DECLARATIONS

### Ethics approval and consent to participate

The Medical Ethics Review Committee (METC) of Amsterdam UMC – Location VU Medical Center waived the review of this study as it was a quality improvement project within regular care (IRB number: IRB00002991; case:19.449). Patients were sent a letter to opt out of the use of their data for this project. The study adheres to "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies".

### Consent for publication

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## COMPETING INTEREST

The authors declare that they have no competing interests.

## FUNDING

This work was funded by the 2019 Amsterdam UMC Innovation Grant: "Het Sepsis Team: Betere Overleving Door Snelle Behandeling" (Project No: 23297).

## AUTHORS' CONTRIBUTIONS

FH, PN, and WJW acquired funding for the study. MS, FH, PN, and WJW conceptualized this study. MS, FH, RV, MR, MD, PN, and WJW actively guided the practical implementation of the sepsis team in the emergency department. MS, RV, and KB collected and curated the data. MS, RV, FH, KB, MR, MD, PN, and WJW collectively investigated the data and decided on the methodology. MS and KB conducted the formal analyses. MS and RV developed resources. FH, PN, and WJW supervised the parts of the research process within their expertise. MS and KB drafted the original manuscript. MS, FH, RV, KB, MR, MD, PN, and WJW reviewed, edited, and agreed with the final version of the manuscript.

## ACKNOWLEDGMENTS

We would like to thank all involved personnel for their efforts in making the sepsis team a success.

## REFERENCES


1. Singer M, Deutschman CS, Seymour CW, et al (2016) The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 315:801
2. van der Poll T, Shankar-Hari M, Wiersinga WJ (2021) The immunology of sepsis. Immunity 54:2450–2464
3. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S (2017) Recognizing Sepsis as a Global Health Priority - A WHO Resolution. N Engl J Med 377:414–417
4. Rudd KE, Johnson SC, Agesa KM, et al (2020) Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet 395:200–211
5. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, Angus DC, Reinhart K (2016) Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. Am J Respir Crit Care Med 193:259–272
6. Marshall JC, Dellinger RP, Levy M (2010) The Surviving Sepsis Campaign: a history and a perspective. Surg Infect (Larchmt) 11:275–281
7. Dellinger RP (2015) The Future of Sepsis Performance Improvement. Crit Care Med 43:1787–1789
8. Evans L, Rhodes A, Alhazzani W, et al (2021) Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. https://doi.org/10.1007/s00134-021-06506-y
9. Levy MM, Rhodes A, Phillips GS, et al (2015) Surviving sepsis campaign: Association between performance metrics and outcomes in a 7.5-year study. Crit Care Med 43:3–12
10. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A (2010) Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: Results from the Portuguese community-acquired sepsis study (SACiUCI study). Crit Care. https://doi.org/10.1186/cc9008
11. Miller RR, Dong L, Nelson NC, Brown SM, Kuttler KG, Probst DR, Allen TL, Clemmer TP (2013) Multicenter implementation of a severe sepsis and septic shock treatment bundle. Am J Respir Crit Care Med 188:77–82
12. Castellanos-Ortega Á, Suberviola B, García-Astudillo LA, Holanda MS, Ortiz F, Llorca J, Delgado-Rodríguez M (2010) Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: Results of a three-year follow-up quasi-experimental study. Crit Care Med 38:1036–1043
13. van Zanten ARH, Brinkman S, Arbous MS, Abu-Hanna A, Levy MM, de Keizer NF (2014) Guideline bundles adherence and mortality in severe sepsis and septic shock. In: Crit Care Med. Lippincott Williams and Wilkins, pp 1890–1898
14. Gray A, Ward K, Lees F, Dewar C, Dickie S, McGuffie C (2013) The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. Emerg Med J 30:397–401
15. Djurkovic S, Baracaldo JC, Guerra JA, Sartorius J, Haupt MT (2010) A survey of clinicians addressing the approach to the management of severe sepsis and septic shock in the United States. https://doi.org/10.1016/j.jcrc.2010.04.005
16. Viale P, Tedeschi S, Scudeller L, et al (2017) Infectious Diseases Team for the Early Management of Severe Sepsis and Septic Shock in the Emergency Department. Clin Infect Dis 65:1253–1259
17. Schinkel M, Nanayakkara PWB, Wiersinga WJ (2022) Sepsis Performance Improvement Programs: From Evidence Toward Clinical Implementation. Critical Care 2022 26:1 26:1–8
18. Damiani E, Donati A, Serafini G, Rinaldi L, Adrario E, Pelaia P, Busani S, Girardis M (2015) Effect of Performance Improvement Programs on Compliance with Sepsis Bundles and Mortality: A Systematic Review and Meta-Analysis of Observational Studies. https://doi.org/10.1371/journal.pone.0125827
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 370:1453–1457

20. Startpagina - Sepsis fase 1 - Richtlijn - Richtlijnendatabase. https://richtlijnendatabase.nl/richtlijn/sepsis_fase_1/startpagina_-_sepsis_fase_1.html. Accessed 6 Jul 2022
21. Dutch patient safety programme (English) - VMSzorg. https://www.vmszorg.nl/vms-veiligheidsprogramma/dutch-patient-safety-programme-english/. Accessed 6 Jul 2022
22. Brown SM, Grissom CK, Moss M, Rice TW, Schoenfeld D, Hou PC, Thompson BT, Brower RG (2016) Nonlinear Imputation of Pao2/Fio2 From Spo2/Fio2 Among Patients With Acute Respiratory Distress Syndrome. Chest 150:307–313
23. Imputed PaO2 from SpO2 Calculator - Open Critical Care. https://opencriticalcare.org/imputed-pao2-calculator/. Accessed 12 Jul 2022
24. R Core Team (2022) R: A Language and Environment for Statistical Computing.
25. Sterling SA, Miller R, Pryor J, Puskarich MA, Jones AE (2015) The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis HHS Public Access. Crit Care Med 43:1907–1915
26. Rothrock SG, Cassidy DD, Barneck M, et al (2020) Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. Ann Emerg Med. https://doi.org/10.1016/j.annemergmed.2020.04.042
27. Alam N, Oskam E, Stassen PM, et al (2018) Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. Lancet Respir Med 6:40–50
28. Schinkel M, Paranjape K, Kundert J, Nannan Panday RS, Alam N, Nanayakkara PWB (2021) Towards Understanding the Effective Use of Antibiotics for Sepsis. Chest. https://doi.org/10.1016/j.chest.2021.04.038
29. Schinkel M, Nannan Panday RS, Joost Wiersinga W, Nanayakkara PWB (2020) Timeliness of antibiotics for patients with sepsis and septic shock. J Thorac Dis 12:S66–S71
30. Simon EL, Truss K, Smalley CM, Mo K, Mangira C, Krizo J, Fertel BS (2022) Improved hospital mortality rates after the implementation of emergency department sepsis teams. Am J Emerg Med 51:218–222
31. Arabi YM, Al-Dorzi HM, Alamry A, et al (2017) The impact of a multifaceted intervention including sepsis electronic alert system and sepsis response team on the outcomes of patients with sepsis and septic shock. Ann Intensive Care 7:1–10
32. Panis LJGG, Verheggen FWSM, Pop P (2002) To stay or not to stay. The assessment of appropriate hospital stay: a Dutch report. Int J Qual Health Care 14:55–67
33. Hoekman LM, Smits MMV, Koolman X (2020) The Dutch COVID-19 approach: Regional differences in a small country. Health Policy Technol 9:613
34. Schnegelsberg A, Mackenhauer J, Pedersen M, Nibro H, Kirkegaard H (2014) Delayed admission to the ICU is associated with increased in-hospital mortality in patients with community-acquired severe sepsis or shock. Crit Care 18:P241
35. Fernando SM, Rochwerg B, Reardon PM, Thavorn K, Seely AJE, Perry JJ, Barnaby DP, Tanuseputro P, Kyeremanteng K (2018) Emergency Department disposition decisions and associated mortality and costs in ICU patients with suspected infection. Crit Care 22:1–10
36. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC (2020) Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA - Journal of the American Medical Association 324:782–793
37. Couarraze S, Delamarre L, Marhar F, et al (2021) The major worldwide stress of healthcare professionals during the first wave of the COVID-19 pandemic – the international COVISTRESS survey. PLoS One 16:e0257840
38. Staat van Infectieziekten in Nederland, 2019 | RIVM. https://www.rivm.nl/publicaties/staat-van-infectieziekten-in-nederland-2019. Accessed 12 Jul 2022
39. Staat van infectieziekten in Nederland, 2020 | RIVM. https://www.rivm.nl/publicaties/staat-van-infectieziekten-in-nederland-2020. Accessed 12 Jul 2022
40. Brueggemann AB, Jansen van Rensburg MJ, Shaw D, et al (2021) Changes in the incidence

of invasive disease due to Streptococcus pneumoniae, Haemophilus influenzae, and Neisseria meningitidis during the COVID-19 pandemic in 26 countries and territories in the Invasive Respiratory Infection Surveillance Initiative: a prospective analysis of surveillance data. Lancet Digit Health 3:e360–e370

# SUPPLEMENTARY APPENDIX

## Tables



**e-Table 1.** Characteristics of patients presenting to the emergency department with a suspected infection and Modified Early Warning Score ≥ 3. A comparison is made between the patients in the postintervention phase of the implementation of the sepsis improvement program in the ED for whom the sepsis team was or was not activated.

| Characteristics | Totals (n=207) | Sepsis team (n=133) | No sepsis team (n=74) | P-value |
|---|---|---|---|---|
| Age, year | 67 (56-76) | 67 (56-76) | 66 (56-74.75) | 0.724 |
| Sex, male | 126 (60.9%) | 83 (62.4%) | 43 (58.1%) | 0.646 |
| Charlson Comorbidity Index | 5 (3-8) | 5 (3-7) | 4 (4-9) | 0.078 |
| MEWS in ED | 5 (3-6) | 5 (4-6) | 4 (3-5) | 0.007 |
| Do not resuscitate orders | 73 (35.3%) | 51 (38.3%) | 22 (29.7%) | 0.140 |
| **Diagnostic workup and treatment** | | | | |
| Complete MEWS recorded | 162 (78.3%) | 113 (85.0%) | 49 (66.2%) | 0.003 |
| Time to MEWS ≥ 3, minutes | 16 (8-30.5) | 16 (8-30) | 16 (7-30.75) | 0.784 |
| Blood culture taken in the ED | 180 (87.0%) | 116 (87.2%) | 64 (86.5%) | 1.000 |
| Time to blood culture, minutes | 28.5 (10-57.25) | 24 (9.75-49.75) | 43.5 (21.5-65.25) | 0.009 |
| Antibiotics administered in the ED | 153 (73.9%) | 106 (79.7%) | 47 (63.5%) | 0.017 |
| Time to first antibiotics, minutes | 75 (43-151) | 66 (40-123.75) | 126 (55-253) | 0.001 |
| Antibiotics administered and blood culture taken in the ED | 139 (67.1%) | 97 (72.9%) | 42 (56.8%) | <0.001 |
| Blood cultures before antibiotics | 123 (88.5%) | 86 (88.7%) | 37 (88.1%) | 0.626 |
| Lactate measurement | 134 (64.7%) | 97 (72.9%) | 37 (50.0%) | 0.002 |
| Repeat measurement of lactate | 51 (38.1%) | 42 (43.3%) | 9 (24.3%) | 0.085 |

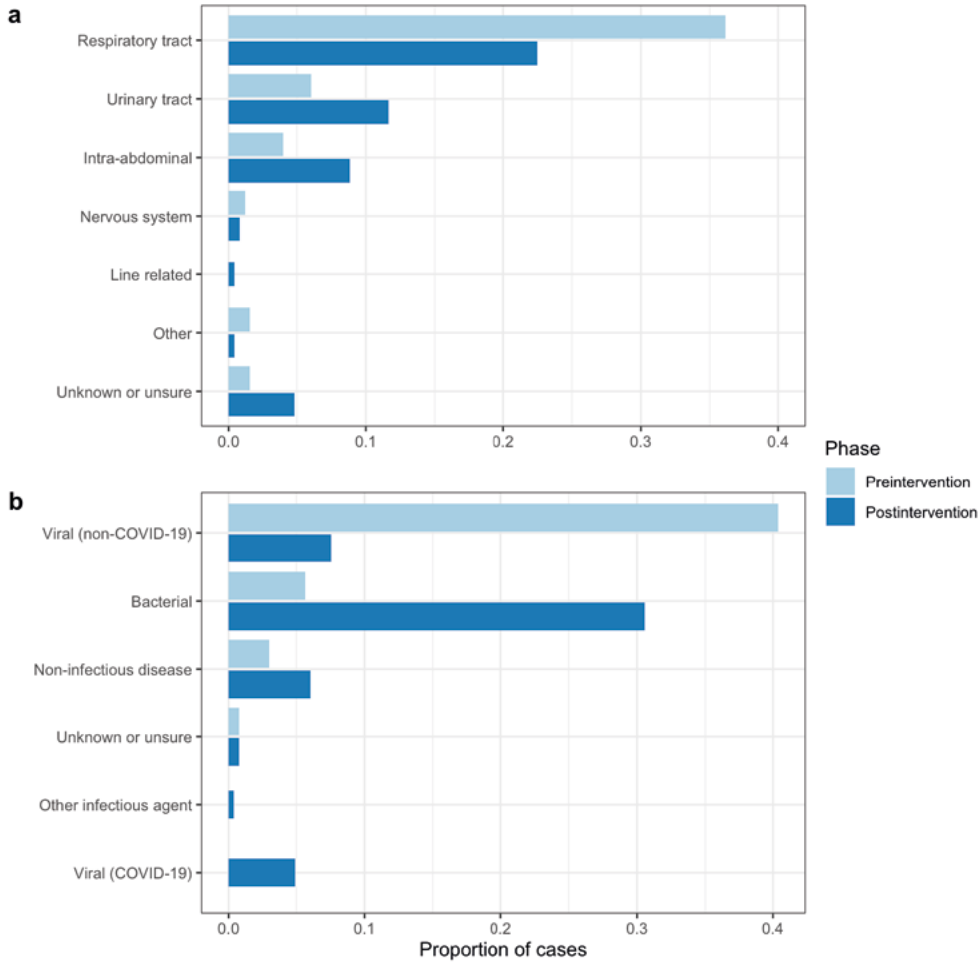
ED = Emergency Department; MEWS = Modified Early Warning Score. Data are presented with no. (%) or median (interquartile range). The results of the indented characteristics are calculated based on the number of patients in the main group of that characteristic.

**e-Table 2.** Outcomes of patients presenting to the emergency department with a suspected infection and Modified Early Warning Score ≥ 3. A comparison is made between the patients in the postintervention phase of the implementation of the sepsis improvement program in the ED for whom the sepsis team was or was not activated.

| Outcome | Totals (n=207) | Sepsis team (n=133) | No sepsis team (n=74) | P-value |
|---|---|---|---|---|
| ED length of stay, minutes | 294 (232-372.50) | 286 (221-407) | 300.5 (256-361.25) | 0.256 |
| Hospital admission | 179 (86.5%) | 117 (88.0%) | 62 (83.8%) | 0.527 |
| Hospital length of stay, days | 5 (3-10) | 5 (3-10) | 5 (3-9) | 0.386 |
| ICU admission | 27 (13.0%) | 22 (16.5%) | 5 (6.8%) | 0.123 |
| Immediate ICU admission from ED | 25 (92.6%) | 22 (100%) | 3 (60.0%) | 0.033 |
| ICU length of stay, days | 2 (1.0-11.0) | 2 (1.25-11.50) | 3 (1-3) | 0.726 |
| 28-day mortality | 25 (12.1%) | 19 (14.3%) | 6 (8.1%) | 0.278 |
| 28-day readmission | 26 (12.6%) | 14 (10.5%) | 12 (16.2%) | 0.532 |
| SOFA score ≥ 2 within 72 hours | 136 (65.7%) | 98 (73.7%) | 38 (51.4%) | 0.002 |

ED = Emergency Department; ICU = Intensive Care Unit; SOFA = Sequential Organ Failure Assessment. Data are presented with no. (%) or median (interquartile range). The results of the indented characteristics are calculated based on the subset of patients in the main group of that characteristic.

## Figures




**e-Figure 1. The suspected sites of infection on presentation (a) and most likely type of infection at hospital discharge (b) of patients presenting to the emergency department with a suspected infection and Modified Early Warning Score ≥ 3.** The bars are stratified by phase of the study (preintervention/ postintervention), which both add up to 1 (100%).

## STOBE Checklist

**STROBE Statement.** checklist of items that should be included in reports of observational studies

| | Item No | Recommendation | Page No |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2 |
| **Introduction** | | | |
| Background/ rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 4 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 5 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 5 |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up | 5 |
| | | *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls | |
| | | *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed | NA |
| | | *Case-control study*—For matched studies, give matching criteria and the number of controls per case | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6 |
| Bias | 9 | Describe any efforts to address potential sources of bias | 7 |
| Study size | 10 | Explain how the study size was arrived at | 6 |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 6 |
| NAStatistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 6, 7 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | 6, 7 |
| | | (c) Explain how missing data were addressed | 6 |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed | NA |
| | | *Case-control study*—If applicable, explain how matching of cases and controls was addressed | |
| | | *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | |
| | | (e) Describe any sensitivity analyses | 6, 7 |

**STROBE Statement.** continued

| | Item No | Recommendation | Page No |
|---|---|---|---|
| **Results** | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | 8 |
| | | (b) Give reasons for non-participation at each stage | 8 |
| | | (c) Consider use of a flow diagram | NA |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | Tab 1 |
| | | (b) Indicate number of participants with missing data for each variable of interest | NA |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) | 8 |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time<br>*Case-control study*—Report numbers in each exposure category, or summary measures of exposure<br>*Cross-sectional study*—Report numbers of outcome events or summary measures | Tab 2 |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 8, 9 |
| | | (*b*) Report category boundaries when continuous variables were categorized | NA |
| | | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | NA |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 9, 10 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 11 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 13 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 12 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 12, 13 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 16 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

# PART II

## THE DIAGNOSTIC WORKUP OF SEPSIS

# CHAPTER 4

## CLINICAL APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN SEPSIS: A NARRATIVE REVIEW

M. Schinkel, MD[1]*, K. Paranjape, Msc[1]*, R.S. Nannan Panday, MD, PhD[1], N. Skyttberg, MD[2], P.W.B. Nanayakkara, MD, PhD, FRCP[1]

[1] Section Acute Medicine, Department of Internal Medicine, Amsterdam UMC, VU University Medical Center. De Boelelaan 1118, 1081HZ, Amsterdam, The Netherlands.
[2] Department of Learning, Informatics, Management and Ethics, Health Informatics Centre, Karolinska Institutet, 171 77, Stockholm, Sweden.
*Both authors contributed equally to the manuscript.

*Computers in Biology and Medicine 115 (2019) 103488*
*https://doi.org/10.1016/j.compbiomed.2019.103488*

## ABSTRACT

Many studies have been published on a variety of clinical applications of artificial intelligence (AI) for sepsis, while there is no overview of the literature. The aim of this review is to give an overview of the literature and thereby identify knowledge gaps and prioritize areas with high priority for further research.

A literature search was conducted in PubMed from inception to February 2019. Search terms related to AI were combined with terms regarding sepsis. Articles were included when they reported an area under the receiver operator characteristics curve (AUROC) as outcome measure.

Fifteen articles on diagnosis of sepsis with AI models were included. The best performing model reached an AUROC of 0.97. There were also seven articles on prognosis, predicting mortality over time with an AUROC of up to 0.895. Finally, there were three articles on assistance of treatment of sepsis, where the use of AI was associated with the lowest mortality rates. Of the articles, twenty-two were judged to be at high risk of bias or had major concerns regarding applicability. This was mostly because predictor variables in these models, such as blood pressure, were also part of the definition of sepsis, which led to overestimation of the performance.

We conclude that AI models have great potential for improving early identification of patients who may benefit from administration of antibiotics. Current AI prediction models to diagnose sepsis are at major risks of bias when the diagnosis criteria are part of the predictor variables in the model. Furthermore, generalizability of these models is poor due to overfitting and a lack of standardized protocols for the construction and validation of the models. Until these problems have been resolved, a large gap remains between the creation of an AI algorithm and its implementation in clinical practice.

## INTRODUCTION

Healthcare today is generating large amounts of data, often dispersed between separate systems[1]. Vital sign monitors, laboratory test results, progress notes and medications along with billing data are stored in electronic medical records[2]. This is a challenge for physicians as they are inundated with so much information, that they first need to collect and understand the data before using it to make a decision. On the other hand, technologies such as Artificial Intelligence (AI) can be applied to gain insights from multiple data sources to enable predictions that can augment the physician's decision-making abilities and improve patient outcomes. AI is a scientific discipline that aims to understand and design computer systems that display intellectual processes, such as reasoning and decision-making, that are otherwise only characteristic of humans [3,4]. For diagnosing conditions, predicting patient outcomes and assisting treatment, Machine Learning has emerged as a popular discipline of AI [5]. Within Machine Learning, Supervised Learning and Reinforcement Learning are being widely used[6]. In Supervised Learning[7], models are trained on known inputs. They output predictions based on evidence in the presence of uncertainty. Reinforcement Learning[8], on the other hand, is the ability to discover which action yields the best outcome through trial and error. Each action affects the next and the user has to plan ahead to select actions that will optimize the outcome. The machine not only considers the immediate effect of certain treatments, but also the long-term benefit to a patient. Complex situations, where multiple and poorly understood mechanisms interact, are perfect areas to implement AI in healthcare, as AI models might be able to identify unforeseen interactions[9]. Sepsis is such an area that is ripe for AI[10].

Sepsis is a life-threatening condition in which early detection and intervention are key in reducing mortality[11]. As per the sepsis-3-criteria[12], sepsis is currently defined as an acute increase in Sequential Organ Failure Assessment (SOFA) score of ≥2 points, indicating life threatening organ dysfunction, due to suspected infection. This is associated with an in-hospital mortality of about 10%[12]. In the early stages of the disease, sepsis is relatively easy to treat with source control and broad spectrum antibiotics[11]. However, diagnosing sepsis in this stage of the disease remains a challenge. In the later stages of the disease, sepsis becomes much easier to diagnose, but extremely hard to treat. With current diagnostic and prognostic tools, it is difficult for physicians to identify patients with sepsis early and to predict their prognoses to decide upon the best treatment strategy for the individual patient. One of the many reasons behind this, is that sepsis is a very heterogeneous syndrome. Patients may develop sepsis based on different pathophysiological mechanisms and may present with different clinical phenotypes[13]. About one in five patients that present to the emergency department with suspected sepsis does not show any signs of organ dysfunction, while they will develop this within 48 hours of admission[14]. Furthermore, bedside screening tools to detect these patients, like the quick Sequential Organ Failure Assessment (qSOFA), lack sensitivity[15,16]. To improve patient outcomes, it is of the essence to improve time to diagnosis and accuracy of the prognosis for patients with sepsis. Some patients who are initially not even categorised as having sepsis might benefit greatly from early administration of antibiotics[17]. AI prediction models, which have shown to be useful for diagnosing and prognostication in other fields of medicine[18,19], could potentially add much value to these areas for patients with sepsis.

In the last decade, a substantial amount of literature has been published on clinical applications of AI for sepsis. The aim of this review is to give an overview of the literature and thereby identify knowledge gaps and prioritize areas with high potential for further research on applications of AI for sepsis. We will focus on AI models that could be valuable in a clinical setting.

# METHODS



## Study design

The aim of this study was to provide an overview of the research field of AI in sepsis. A narrative review was considered the most appropriate approach, as it has been considered appropriate to "tell the story" of the evidence. Narrative reviews are described as a good choice in situations when there are disparate interventions or when there is dissimilarity of outcome measures and follow-up times in the analysed material[20].

## Study identification/search strategy

A literature search was conducted in the bibliographic database PubMed from inception to February 2019. Search terms related to AI were combined with terms regarding sepsis (See appendix for further details). Additional articles were included based on expert opinion.

## Study selection

Articles were screened by title and abstract by two reviewers (KP and MS). Studies were selected when types of AI, such as artificial neural networks, random forest models or gradient-boosted tree models were used in patients with sepsis. Logistic regression models are widely used in medical literature for statistical analysis, but rarely for predictive models. Therefore, logistic regression was not included as a type of AI for this particular review. Once selected, full texts were appraised. Articles were included when an area under the receiver operator characteristics curve (AUROC) was reported as outcome for diagnosis or prognosis of sepsis. AUROC was chosen because it is robust to differences in the prevalence of the outcomes in the various studies. Articles regarding assistance of treatment in sepsis were also included when a difference in outcome was reported by means. When full texts were not freely available, the article was requested from the VU Amsterdam Medical Center library. 3 articles that were conference abstracts were excluded. Articles were also excluded when there was no link to clinical practice, which was the case in articles that, for example, used AI to extract information from genes[21]. Systematic reviews were also excluded (See Figure 1 for further details).

## Categories

The selected articles were categorized into three groups, to give an overview of the different areas of applications of AI for sepsis: diagnosis, prognosis and treatment. A subcategory was added to the diagnosis section: articles on predictions regarding the pathogens causing sepsis.

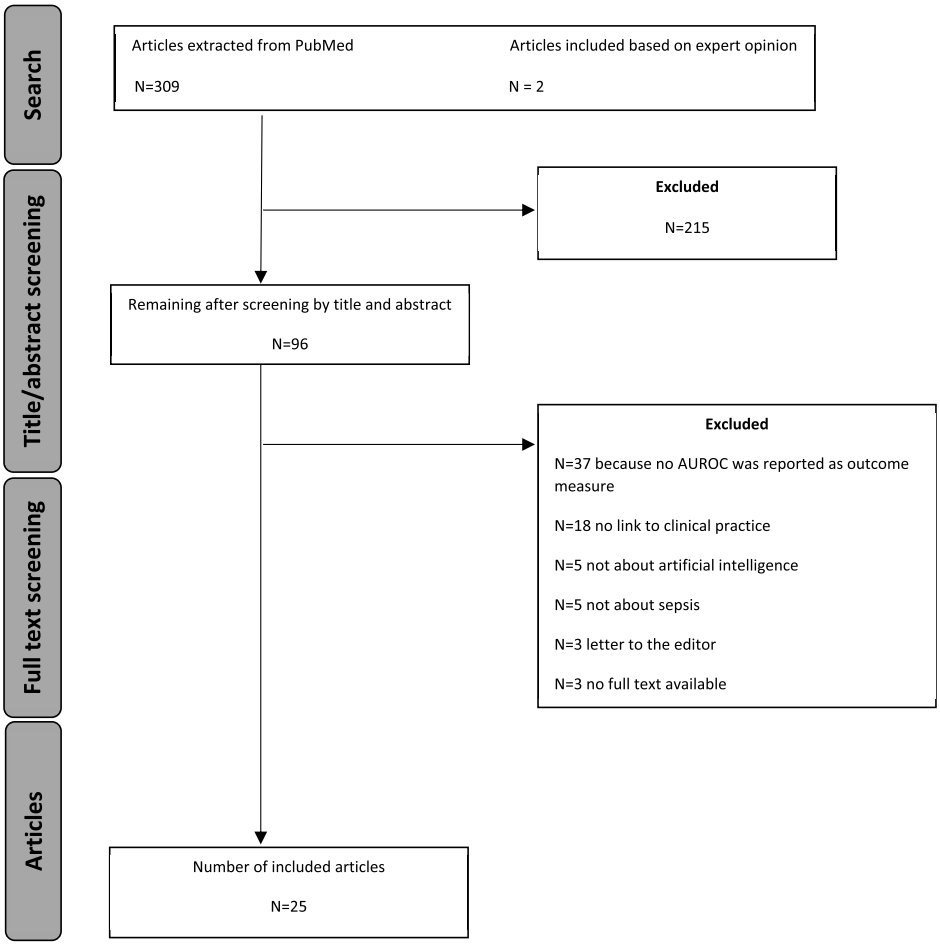



**Figure 1. Flowchart of article selection.**

### Study quality assessment

The risk of bias and concerns regarding the applicability of the included studies was examined using the recently developed PROBAST-tool, which was specifically designed to assess these qualities in studies on prediction models[22]. The PROBAST-tool focuses on four domains: participant selection, predictor variables, outcomes and statistical analysis. The questions within these domains address frequently encountered problems, such as the lack of available data at the time when a model should be used.

## RESULTS

### Characteristics

The search, supplemented with two articles based on expert opinion, yielded 311 articles. After screening by title and abstract, ninety-six were selected, as they reported on some application of

AI in patients with sepsis. After full texts were appraised, twenty-two articles[10,23–42] on diagnosis or prognosis of sepsis through AI were identified that reported an AUROC as outcome measure. Another three articles on assisting treatment of sepsis through AI were included[43–45]. The characteristics of the twenty-five included studies are presented in table 1. Table 3 elaborates on the specific types of AI models in these studies.



## Study quality

The risk of bias and concerns regarding the applicability of the studies was examined using the PROBAST-tool. Two studies were found to be at low risk of bias and low concern regarding applicability, while twenty-two studies were found to be at high risk of bias or high concern regarding applicability. One article could not be assessed with the PROBAST-tool as the development of the model was not described[44] (See Table 2 for further details on the PROBAST assessments).

## Diagnosis

Of the included articles, eleven reported on diagnosing sepsis (See table 1 for details on study populations). Barton and colleagues created an algorithm that predicted sepsis onset 48-hours in advance with an AUROC of 0.83, using just vital signs. Delahanty et al. created a model to predict the onset of sepsis during hospital admission, according to sepsis criteria as proposed by Rhee and colleagues [23,46]. This Risk of Sepsis score (RoS) reached an AUROC of 0.93 in the first hour of admission and increased to 0.97 after 24 hours. Desautels and colleagues created an algorithm (InSight) to predict sepsis onset in an intensive care unit (ICU) population[34]. The model used vital signs and age and reached an AUROC of 0.880. Mao et al. validated the InSight algorithm in a different ICU dataset and detected sepsis 4 hours before onset with an AUROC of 0.92[38]. Also, 4 hours before onset, the algorithm predicted septic shock with an AUROC of 0.96. Kam and colleagues created a model to predict sepsis with an AUROC of 0.929[37]. Kaji and colleagues predicted same-day and next-day sepsis[36]. Same-day sepsis onset prediction models achieved an AUROC of 0.952, while this was 0.876 for prediction of next-day sepsis. Nemati et al. reported on an algorithm with an AUROC of 0.85 to predict sepsis 4 hours before onset[39]. Saqib et al. reported on a model that predicted sepsis in an ICU population with an AUROC of 0.696 [42]. Shashikumar et al. predicted sepsis in an ICU population 4 hours in advance with an AUROC of 0.78 [39].Taneja et al. predicted sepsis onset with a model based on vital parameters, as well as individual biomarkers[40]. The AUROC was 0.81. Henry and colleagues created a real-time warning score to predict the onset of sepsis a median of 28.2 hours before onset with an AUROC of 0.83 [41].

Four articles reported on predictions regarding the pathogens that caused sepsis. Van Steenkiste and colleagues used an AI model to predict positive blood cultures[25]. The AUROC was 0.98 when 72 hours of data was used for the prediction, while Ratzinger and colleagues predicted bacteraemia with an AUROC of 0.73 at the moment the blood cultures were drawn [28]. In the study by Oonsivalai et al., the best model to predict whether a pathogen was susceptible to certain antibiotics reached an AUROC of 0.80, for predicting susceptibility to ceftriaxone[26]. Lamping and colleagues used an AI model to distinguish sepsis from non-infectious SIRS in critically ill children, achieving an AUROC of 0.78 [27].

**Table 1.** Characteristics of the included studies.

| Author, year | Study design | Setting | Database (MIMIC = Medical Information Mart for Intensive Care) | No. predictor variables in model | Outcome | PROBAST-assessment (Risk of bias; concern with applicability) |
|---|---|---|---|---|---|---|
| **Diagnosis** | | | | | | |
| **Delahanty, 2019** | Retrospective | Emergency Department | Hospital database (2.759.529 patient encounters) | 13 | AUROC: 0.93 at 1-hour, AUROC 0.97 at 24-hours | ROB: high, applicability: high |
| **Desautels, 2016** | Retrospective | Intensive Care | MIMIC-III | 8 | AUROC: 0.880 at disease onset | ROB: high, applicability: low |
| **Kaji, 2019** | Retrospective | Intensive Care | MIMIC-III | 119 | AUROC: 0.952 at same-day, 0.876 at next-day | ROB: high, applicability: unclear |
| **Kam, 2017** | Retrospective | Intensive Care | MIMIC-III | 9 | AUROC: 0.929 | ROB: high, applicability: low |
| **Mao, 2018** | Retrospective | Hospital wide | Hospital database (17.467.987 patient encounters) MIMIC-III | 6 | AUROC: 0.92 4-hours before sepsis onset. | ROB: high, applicability: low |
| **Nemati, 2017** | Retrospective | Intensive Care | Hospital database (27.527 patient encounters) MIMIC-III | 65 | AUROC: 0.85 4-hours before sepsis | ROB: high, applicability: high |
| **Taneja, 2017** | Retrospective | Hospital wide | Hospital database (444 patient encounters) | 21 | AUROC: 0.81 at disease onset | ROB: high , applicability: high |
| **Henry, 2015** | Retrospective | Intensive care | MIMIC-III | 26 | AUROC: 0.83 28.2-hours before sepsis onset. | ROB: high , applicability: high |
| **Saqib, 2018** | Retrospective | Intensive care | MIMIC-III | 12 | AUROC: 0.696 | ROB: high, applicability: low |
| **Shashikumar, 2017** | Retrospective | Intensive Care | Hospital database (242 patient encounters) | Unclear | AUROC: 0.78 4-hours before sepsis onset. | ROB: high, applicability: low |
| **Barton, 2019** | Retrospective | Hospital Wide | Hospital database (91,445 patient encounters) MIMIC-III | 6 | AUROC: 0.83 48-hours before onset. | ROB: high, applicability: low |

**Table 1.** continued

| Author, year | Study design | Setting | Database (MIMIC = Medical Information Mart for Intensive Care) | No. predictor variables in model | Outcome | PROBAST-assessment (Risk of bias; concern with applicability) |
|---|---|---|---|---|---|---|
| **Pathogen prediction** | | | | | | |
| **Van Steenkiste, 2018** | Retrospective | Hospital wide | Hospital database (2177 patient encounters) | 9 | AUROC: 0.99 with 72 hours of data | ROB: low, applicability: low |
| **Oonsivalai, 2018** | Retrospective | Hospital wide | Hospital database (243 patient encounters) | 35 | AUROC: 0.80 for ceftriaxone susceptibility | ROB: high, applicability: high |
| **Lamping, 2018** | Prospective, RCT | Pediatric ICU | Hospital based (230 patient encounters) | 8 | AUROC: 0.78 for infectious vs. non-infectious SIRS | ROB: high, applicability: high |
| **Ratzinger, 2018** | Prospective | Hospital wide | Hospital based (466 patient encounters) | 21 | AUROC: 0.73 for bacteraemia. | ROB: high, applicability: low |
| **Prognosis** | | | | | | |
| **Aushev, 2018** | Retrospective | Intensive care | ShockOmics | 80 | AUROC: 0.845 for ICU mortality | ROB: high, applicability: high |
| **Dybowski, 1996** | Retrospective | Intensive care | Hospital database (4484 patient encounters) | 11 | AUROC: 0.863 for in-hospital mortality | ROB: high, applicability: high |
| **Garcia-Gallo, 2018** | Retrospective | Intensive care | MIMIC-III | 18 | AUROC: 0.8083 for 1 year mortality | ROB: high, applicability: low |
| **Jaimes, 2005** | Retrospective | Emergency department | Hospital database (542 patient encounters) | 10 | AUROC: 0.8782 for 28-day mortality | ROB: low, applicability: low |
| **Meiring, 2018** | Retrospective | Intensive care | MIMIC-II<br>MIMIC-III | 25 | AUROC: 0.895 for mortality at ICU discharge | ROB: low, applicability: low |
| **Taylor, 2016** | Retrospective | Emergency department | Hospital database (4676 patient encounters) | 25 | AUROC: 0.86 for in-hospital mortality | ROB: high, applicability: high |
| **Ward, 2017** | Retrospective | Trials/Studies | Hospital database (2514 patient encounters) | 18 | AUROC: 0.79 for 30-day mortality | ROB: high, applicability: high |

**Table 1.** continued

| Author, year | Study design | Setting | Database (MIMIC = Medical Information Mart for Intensive Care) | No. predictor variables in model | Outcome | PROBAST-assessment (Risk of bias; concern with applicability) |
|---|---|---|---|---|---|---|
| **Treatment assistance** | | | | | | |
| **Komorowski, 2018** | off-policy evaluation | Intensive care | MIMIC-III<br>eICU | 48 | AI policy associated with lowest mortality | ROB: high, applicability: high |
| **Merouani, 2008** | Prospective, randomized | Intensive care | Hospital database (42 patient encounters) | 2 | Median duration of shock significantly shorter (28.5 hours versus 57.5 hours). | ROB: -, applicability: - |
| **Shimbukuro, 2017** | Randomized controlled trial | Intensive care | Hospital database (142 patient encounters) | 8 | In-hospital mortality decreased by 12.4 percentage points | ROB: high, applicability: low |

**Table 2.** Detailed PROBAST-assessments of the included studies.

| | Risk of bias (ROB) | | | | Applicability | | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| Study | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome | ROB | Applicability |
| **Diagnosis** | | | | | | | | | |
| Delahanty, 2019 | + | - | - | + | + | + | - | - | - |
| Desautels, 2016 | - | - | - | - | + | + | + | - | + |
| Kaji, 2019 | + | ? | - | - | + | ? | + | - | ? |
| Kam, 2017 | - | + | - | - | + | + | + | - | + |
| Mao, 2018 | - | - | - | + | + | + | + | - | + |
| Nemati, 2017 | - | - | - | - | + | - | + | - | - |
| Taneja, 2011 | - | - | - | - | - | - | - | - | - |
| Henry, 2015 | + | - | - | - | + | - | + | - | - |
| Saqib, 2018 | - | - | + | + | - | - | - | + | - |
| Shashikumar, 2017 | + | + | - | - | + | + | + | - | + |
| Barton, 2019 | + | + | - | + | + | + | + | - | + |
| **Pathogen prediction** | | | | | | | | | |
| Van Steenkiste, 2018 | + | + | + | + | + | + | + | + | + |
| Oonsivalai, 2018 | + | - | + | - | + | - | + | - | - |
| Lamping, 2018 | + | + | + | - | - | + | + | - | - |
| Ratzinger, 2018 | + | + | + | - | + | + | + | - | + |
| **Prognosis** | | | | | | | | | |
| Aushev, 2018 | - | - | + | - | + | - | + | - | - |
| Dybowski, 1996 | - | + | + | - | - | + | + | - | - |
| Garcia-Gallo, 2018 | - | + | + | + | + | + | + | - | + |
| Jaimes, 2005 | + | + | + | + | + | + | + | + | + |
| Meiring, 2018 | + | + | + | + | + | + | + | + | + |
| Taylor, 2016 | - | - | + | - | + | - | + | - | - |
| Ward, 2017 | - | + | + | + | - | + | + | - | - |
| **Treatment assistance** | | | | | | | | | |
| Komorowski, 2018 | - | - | + | + | + | - | + | - | - |
| Merouani, 2008 | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| Shimabukuro, 2017 | - | - | - | - | + | + | + | - | + |

PROBAST = Prediction model Risk Of Bias ASsessment Tool; ROB = risk of bias. * + indicates low ROB/low concern regarding applicability; – indicates high ROB/high concern regarding applicability; and ? indicates unclear ROB/unclear concern regarding applicability.

**Table 3.** Specific types of artificial intelligence models

| Author | Year | Type of Learning | Type of model |
|---|---|---|---|
| **Delahanty** | 2019 | Supervised | Gradient-boosted tree model |
| **Desautels** | 2016 | Supervised | Gradient-boosted tree model |
| **Mao** | 2018 | Supervised | Gradient-boosted tree model |
| **Kam** | 2017 | Reinforced | Long short-term memory |
| **Kaji** | 2019 | Reinforced | Neural Network |
| **Nemati** | 2018 | Supervised | Modified Weilbull-Cox proportional hazards model |
| **Taneja** | 2017 | Supervised | Support Vector Machine |
| **Van Steenkiste** | 2018 | Reinforced | Long short-term memory neural network |
| **Oonsivalai** | 2018 | Supervised | Random Forest Model |
| **Dybowski** | 1996 | Reinforced | Artificial Neural Network |
| **Taylor** | 2016 | Supervised | Random Forest model |
| **Aushev** | 2018 | Supervised | Machine Learning |
| **Meiring** | 2018 | Reinforced | Deep Learning Model |
| **Jaimes** | 2005 | Reinforced | Artificial Neural Network |
| **Garcia-Gallo** | 2018 | Supervised | Stochastic Gradient Boosting |
| **Komorowski** | 2018 | Reinforced | Markov decision process |
| **Merouani** | 2008 | Reinforced | Fuzzy Logic |
| **Shimbukuro** | 2017 | Supervised | Machine learning |
| **Henry** | 2015 | Supervised | Cox proportional hazards model |
| **Ward** | 2017 | Supervised | Causal Probabilistic Network |
| **Lamping** | 2018 | Supervised | Random Forest Model |
| **Ratzinger** | 2018 | Supervised | Random Forest Model |
| **Saqib** | 2018 | Supervised | Random Forest Model |
| **Shashikumar** | 2017 | Supervised | Elastic Net logistic classifier |
| **Barton** | 2019 | Supervised | Gradient-boosted tree model |

## Prognosis

Seven studies were included that used AI to predict the outcome of patients with sepsis. Dybowski et al. created an algorithm to predict in-hospital mortality in patients with sepsis[10]. The model reached an AUROC of 0.863. Taylor and colleagues reported on a model to predict in-hospital mortality with an AUROC of 0.86[33]. Furthermore, Aushev et al, reported on a model predicting in-hospital mortality with an AUROC of 0.845[29]. Meiring et al. used an algorithm to predict mortality over time in the ICU[32]. The model reached an AUROC of 0.895. The article by Jaimes et al. described a model that predicted 28-day mortality[31]. The AUROC for this model was 0.8782. Garcia-Gallo et al. aimed to predict 1-year mortality with a model that achieved an AUROC of 0.8039[30]. Ward and colleagues predicted 30-day mortality for patients with an infection or sepsis, reaching an AUROC of 0.79 [35].

## Treatment

We identified three articles regarding assistance of treatment of sepsis using AI. Komorowski et al. created an "artificial intelligent clinician" using reinforcement learning[43]. The aim was to create an algorithm that assisted clinicians by suggesting the best treatment at the right time. The model

was built based on data from two large ICU databases, Medical Information Mart for Intensive Care (MIMIC)-III and eICU Research Institute Database, that are available online. The "AI clinician" suggested doses of intravenous fluids and vasopressors. On average, the AI recommended higher doses of vasopressors and lower doses of fluids when compared to clinicians. The AI suggested doses correlated with the lowest risk of mortality.

Merouani and colleagues used algorithms to improve the weaning rate of vasopressors[44]. The suggestions from the AI model were compared to the clinicians. The duration of septic shock was significantly shorter in the AI group versus the control group (median time in hours: 28.5 versus 57.5; $p < 0.001$). Also, the total amount of vasopressors was reduced significantly (0.6 µg/kg versus 1.4 µg/kg; $p < 0.01$). No significant difference in mortality was observed.

Shimabukuro et al. used the InSight model, which was described in the diagnosis section of our results, and compared it to standard care[45]. The model was trained to generate an alert message to the nurse when the algorithm predicted deterioration of clinical condition to a state of severe sepsis. This would result in a different course of treatment, according to the hospital guidelines. Use of the InSight model resulted in a decrease in in-hospital mortality from 21.3% to 8.96% ($p=0.018$). Furthermore, length of stay in the hospital was reduced from 13.0 to 10.3 days ($p=0.042$) (see Table 3 for details on AI models).

# DISCUSSION

## Diagnosis

Most included studies reported on AI models that predict whether a patient has sepsis or will develop it over time. Diagnosing sepsis in the early stages of the disease remains a challenge because of the complex pathophysiology, heterogeneity and lack of accurate diagnostic tools[15,16]. As early administration of antibiotics might benefit certain patients[14], AI prediction tools have the potential to improve patient outcomes.

We reported on eleven different models that predict sepsis with an AUROC of 0.696 to 0.952, mostly in emergency department and ICU populations (see Table 1). These models outperform current tools for detecting sepsis such as the Sequential Organ Failure Assessment (SOFA), Systemic Inflammatory Response Syndrome criteria (SIRS) and Modified Early Warning score (MEWS). The InSight algorithm achieved an AUROC of 0.880, while this was significantly lower for the SOFA (0.725), SIRS (0.609) and MEWS (0.803) in the same population[45,47]. Our findings are in accordance with a recent meta-analysis by Islam et al. which investigated studies that only reported on AI algorithms to diagnose sepsis in early stages of the disease[48].

As illustrated throughout this review, AI can be applied to generate insights from various data sources. Algorithms use clinical features, laboratory features, patient history, demographics and clinical context to predict the desired outcome measures[49]. In addition, real-time data streams are increasingly being used[50]. The value of AI models, especially when they are based on vital parameters only, is their instant usability. An algorithm can raise alerts in cases where clinicians have not yet thought of sepsis as the diagnosis. The use of laboratory tests, as needed for the SOFA score, or use of upcoming biomarkers to detect sepsis, such as procalcitonin[51], requires an active

decision to test by the clinician. Furthermore, some laboratory tests can take several hours and delay treatment. Some of the algorithms, like the model by Mao et al, can predict sepsis onset 4 hours in advance[38]. These additional hours could be crucial in optimizing treatment, as some patients might benefit from early administration of antibiotics[14]. The same problem arises with blood cultures, which are used to determine the best choice of antibiotics. Results take up to 4 days and can delay optimal treatment[52]. We reported on four articles that used AI to make predictions about the pathogens that caused sepsis. When these algorithms could be used to choose the best treatment before blood culture results are available, the patient outcomes might improve due to early administration of antibiotics.



One issue with the use of AI to diagnose sepsis is that most of the models included in the study are based on either ICU or emergency department patient population and use variables that are commonly measured in these settings (see Table 1). Several studies used the same database to create their algorithms: the Medical Information Mart for Intensive Care (MIMIC) database, which is a large, single-center database that is freely available for research[53]. Using these algorithms in other departments would likely result in decreased accuracy. To be able to use algorithms to capture patients at risk of sepsis across the entire hospital, different models are required.

## Prognosis

We included seven articles reporting on AI based prognostic models for sepsis outcomes. The articles all focused on predicting mortality, at different points in time, for ICU and emergency department populations. We reported on models with AUROC values of 0.79 to 0.90. These values are comparable to the APACHE-II score, which is widely used, with an AUROC of 0.83[54]. Some patients who are initially not even categorized as having sepsis, might decline rapidly and have a high chance of mortality[14]. These patients could benefit remarkably from administration of antibiotics. An AI algorithm that could predict these high mortality rates for certain patients, would therefore be very valuable to clinicians. As shown, these algorithms exist, but they only just outperform current standards. Further optimization of these algorithms could potentially add much value to clinical practice. Notably, none of the studies in the diagnosis or prognosis categories assessed whether the predictions led to more favorable outcomes.

## Assistance of treatment

We identified three studies that focused on using AI to optimally treat patients with sepsis. These AI models were shown to decrease mortality or duration of shock in patients with sepsis. Treatment of patients with sepsis is relatively easy in the early stages of the disease, but becomes much more difficult in the later stages, especially when patients develop septic shock. Consequently, all of the AI algorithms that assist choice of treatment were based on ICU populations. This is where we would expect the biggest impact of using AI for treatment assistance. However, for the general sepsis population research focused exclusively on AI algorithms that assist treatment choice would most likely not add much value to clinical practice.

## PROBAST assessments

To assess the risk of bias and problems with applicability of predictive models in clinical practice, the PROBAST-tool was developed[22]. We used this tool to assess the quality of the included studies (see Table 2). We reported that twenty-two of the articles had either a high risk of bias or major concern regarding applicability, while just two articles had low risk of bias and no concerns regarding applicability. The study by Merouani et al. could not be evaluated, as the development of the model was not described[44]. Several problems are observed frequently and are discussed further. First, as the definition of sepsis or detection of organ dysfunction includes many variables, such as blood pressure and creatinine levels, these often overlap with predictor variables that are used in the AI models. As stated by PROBAST: "If a predictor in the model forms part of the definition or assessment of the outcome that the model predicts, the association between the predictor and outcome will likely be overestimated and estimates of the model performance will be optimistic"[22]. All the included diagnostic models were therefore at high risk of bias in our assessments. A model's accuracy can only truly be assessed when predictors that are in the SOFA-score are not used in the model. This would decrease the accuracy of the models, as these variables are by definition signs of sepsis. As long as the definition of sepsis remains based on clinical parameters, predicting the onset of sepsis with these same parameters will continue to be open to bias. The question arises whether the overestimation matters when the algorithms outperform the current standards. We will not know the true accuracy of these algorithms this way, but leaving out valuable signs of sepsis seems contra-intuitive. The high AUROC values in some of the included studies, such as the AUROC of 0.97 in the study by Delahanty and colleagues [23], could also be explained by overfitting. Overfitting occurs when the algorithm is trained too specifically to predict the outcomes in a particular study population. The algorithm can take into account factors that are normally not associated with the outcome, but do improve accuracy in this particular population. These high AUROC values will likely not be reached when the algorithm is used in a different population of patients. So, it remains questionable whether the high-performance algorithms that we have examined in this narrative review actually outperform current standards in practice. This problem can be addressed by mandatory external validation when such an algorithm is developed.

A second issue, highlighted by the PROBAST assessments, is that most models were built on databases with many missing values. One such databases is the MIMIC-III database, that was used for several of the included studies. Most variables with missing values were excluded from being predictor variables in the studies included here. Even when a dataset is complete, there can be selection bias or confounding factors[55,56]. Thus, variables with a high predictive accuracy might be missed or misinterpreted in the included studies. When predictor variables for the model by Dybowski and colleagues were selected through different statistical methods, just two predictor variables were shared[10]. Since there is little guidance as to how models should be constructed and validated, algorithms that are based on the same dataset, can be very different. This means that the AI models typically have poor generalizability. Different hospitals or departments need to have their own version of a certain model. Standardized protocols for implementing AI in healthcare are therefore a necessity.

The last concern that was raised by the PROBAST assessments is regarding the applicability of these models. Most models use large amounts of predictor variables. Many are not routinely measured. Even when they are measured, it would still be questionable whether the data is available at the right time[57]. When algorithms are used in clinical practice, poor availability of the data would decrease the accuracy. This problem, along with the likelihood of overfitting and poor generalizability, causes a large gap between creating a model in a retrospective database and implementing the model in a clinical setting.



### Nonincluded articles

We did not include papers on diagnosing sepsis with AI that did not report an AUROC as outcome measures. Consequently, no articles with algorithms that used streams of physiologic data to detect sepsis early were included. We would argue that the use of these routinely measured physiomarkers could yield good results since the sepsis criteria today are largely based on physiologic data [12]. Here are 3 papers that addressed this topic. In 2018, Kamaleswaran et al. reported on an AI model that used continuous minute-by-minute physiologic data to predict severe sepsis in children [58]. Depending on the number of hours of data that was used, a sensitivity of up to 76% could be achieved with a specificity of 81%. From the same group, van Wyk and colleagues published two additional papers that reported on AI algorithms that used continuous streams of physiologic data to predict sepsis [59,60]. The first being able to predict sepsis half-hour before onset with an accuracy of 79% [59], while the second predicted sepsis on average 205 minutes earlier than what SIRS criteria would have predicted [60]. So, there is a lot of potential for AI models based on data streams since physiologic data is readily available. But we believe that the problems we have encountered throughout this review, are likely to influence these models as well.

### Strengths and limitations

This article was written by medical professionals, as well as computer and data science experts. This combination of expertise enabled us to highlight essential aspects from all fields.

Despite this strength, there are some limitations. As this is a narrative review, not all available literature on this subject was discussed. State-of-the-art AI techniques such as clustering sepsis into different phenotypes[13], was not discussed. These kinds of projects do not yet translate into clinical practice and are mostly used in research settings. As the aim of this study was to give an overview of possible clinical applications of AI in sepsis, we chose this particular study design.

## CONCLUSION

In early stages of the disease, sepsis is easy to treat, but hard to diagnose. In later stages, sepsis becomes much easier to diagnose, but very hard to treat. AI models have great potential for improving early identification of patients who may benefit from administration of antibiotics. Some AI prediction models seem to outperform current diagnostic tools by a fair margin, but there are many problems with these models, such as the fact that predictor variables like blood pressure, are also part of the current definition of sepsis. This leads to overestimation of the performance of these

AI models. Furthermore, generalizability of these models is very poor. Until these problems have been resolved, a large gap remains between the creation of an AI algorithm and its implementation in clinical practice.

## FINANCIAL SUPPORT

No financial support was received for this work.



## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCE LIST


1. Purkayastha S, Gichoya JW, Siva ;, Addepally A. Implementation of a single sign-on system between practice, research and learning systems. 2017. doi:10.4338/ACI-2016-10-CR-0171
2. Hazelzet JA. Can fuzzy logic make things more clear? *Crit Care*. 2009;13(1):116. https://www.ncbi.nlm.nih.gov/pubmed/19291252.
3. Bali J, Garg R, Bali R. Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? *Indian J Ophthalmol*. 2019;67(1):3-6. doi:10.4103/ijo.IJO_1292_18
4. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8(2). doi:10.7189/jogh.08.020303
5. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2(10):719-731. doi:10.1038/s41551-018-0305-z
6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246. doi:10.1093/bib/bbx044
7. Mahapatra D, Schueffler P, Tielbeek JAW, Buhmann JM, Vos FM. A Supervised Learning Based Approach to Detect Crohn's Disease in Abdominal MR Volumes. In: Springer, Berlin, Heidelberg; 2012:97-106. doi:10.1007/978-3-642-33612-6_11
8. Gottesman O, Johansson F, Komorowski M, et al. Guidelines for reinforcement learning in healthcare. *Nat Med*. 2019;25(1):16-18. doi:10.1038/s41591-018-0310-5
9. Clermont G. Artificial neural networks as prediction tools in the critically ill. *Crit Care*. 2005;9(2):153-154. https://www.ncbi.nlm.nih.gov/pubmed/15774070.
10. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet (London, England)*. 1996;347(9009):1146-1150. https://www.ncbi.nlm.nih.gov/pubmed/8609749.
11. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign. *Crit Care Med*. 2013;41(2):580-637. doi:10.1097/CCM.0b013e31827e83af
12. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801. doi:10.1001/jama.2016.0287
13. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019;321(20):2003. doi:10.1001/jama.2019.5791
14. Quinten VM, van Meurs M, Ligtenberg JJ, Ter Maaten JC. Prehospital antibiotics for sepsis: beyond mortality? *Lancet Respir Med*. 2018;6(3):e8. doi:10.1016/S2213-2600(18)30061-4
15. Goulden R, Hoyle M-C, Monis J, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg Med J*. 2018;35(6):345-349. doi:10.1136/emermed-2017-207120
16. Anand V, Zhang Z, Kadri SS, Klompas M, Rhee C, CDC Prevention Epicenters Program. Epidemiology of qSOFA Criteria in Undifferentiated Patients and Association with Suspected Infection and Sepsis. *Chest*. 2019;0(0). doi:10.1016/j.chest.2019.03.032
17. Quinten VM, van Meurs M, Wolffensperger AE, Ter Maaten JC, Ligtenberg JJM. Sepsis patients in the emergency department: stratification using the Clinical Impression Score, Predisposition, Infection, Response and Organ dysfunction score or quick Sequential Organ Failure Assessment score? *Eur J Emerg Med*. 2018;25(5):328-334. doi:10.1097/MEJ.0000000000000460
18. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22):2402. doi:10.1001/jama.2016.17216
19. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. doi:10.1038/nature21056
20. Melendez-Torres GJ, O'Mara-Eves A, Thomas J, Brunton G, Caird J, Petticrew M. Interpretive analysis of 85 systematic reviews suggests that narrative syntheses and meta-analyses are

incommensurate in argumentation. *Res Synth Methods*. 2017;8(1):109-118. doi:10.1002/jrsm.1231

21. Lu X, Xue L, Sun W, Ye J, Zhu Z, Mei H. Identification of key pathogenic genes of sepsis based on the Gene Expression Omnibus database. *Mol Med Rep*. 2018;17(2):3042-3054. https://www.ncbi.nlm.nih.gov/pubmed/29257295.

22. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration Annals of Internal Medicine RESEARCH AND REPORTING METHODS. *Ann Intern Med*. 2019;170:1-33. doi:10.7326/M18-1377

23. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Ann Emerg Med*. 2019. https://www.ncbi.nlm.nih.gov/pubmed/30661855.

24. Barton C, Chettipally U, Zhou Y, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med*. 2019;109:79-84. doi:10.1016/j.compbiomed.2019.04.027

25. Van Steenkiste T, Ruyssinck J, De Baets L, et al. Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. *Artif Intell Med*. 2018. https://www.ncbi.nlm.nih.gov/pubmed/30420241.

26. Oonsivilai M, Mo Y, Luangasanatip N, et al. Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia. *Wellcome open Res*. 2018;3:131. https://www.ncbi.nlm.nih.gov/pubmed/30756093.

27. Lamping F, Jack T, Rübsamen N, et al. Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children - a data-driven approach using machine-learning algorithms. *BMC Pediatr*. 2018;18(1):112. https://www.ncbi.nlm.nih.gov/pubmed/29544449.

28. Ratzinger F, Haslacher H, Perkmann T, et al. Machine learning for fast identification of bacteraemia in SIRS patients treated on standard care wards: a cohort study. *Sci Rep*. 2018;8(1):12233. https://www.ncbi.nlm.nih.gov/pubmed/30111827.

29. Aushev A, Ripoll VR, Vellido A, et al. Feature selection for the accurate prediction of septic and cardiogenic shock ICU mortality in the acute phase. *PLoS One*. 2018;13(11):e0199089. https://www.ncbi.nlm.nih.gov/pubmed/30457997.

30. García-Gallo JE, Fonseca-Ruiz NJ, Celi LA, Duitama-Muñoz JF. A machine learning-based model for 1-year mortality prediction in patients admitted to an Intensive Care Unit with a diagnosis of sepsis. *Med intensiva*. 2018. https://www.ncbi.nlm.nih.gov/pubmed/30245121.

31. Jaimes F, Farbiarz J, Alvarez D, Martínez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care*. 2005;9(2):R150-6. https://www.ncbi.nlm.nih.gov/pubmed/15774048.

32. Meiring C, Dixit A, Harris S, et al. Optimal intensive care outcome prediction over time using machine learning. *PLoS One*. 2018;13(11):e0206862. https://www.ncbi.nlm.nih.gov/pubmed/30427913.

33. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Acad Emerg Med*. 2016;23(3):269-278. https://www.ncbi.nlm.nih.gov/pubmed/26679719.

34. Desautels T, Calvert J, Hoffman J, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med informatics*. 2016;4(3):e28. https://www.ncbi.nlm.nih.gov/pubmed/27694098.

35. Ward L, Paul M, Andreassen S. Automatic learning of mortality in a CPN model of the systemic inflammatory response syndrome. *Math Biosci*. 2017;284:12-20. https://www.ncbi.nlm.nih.gov/pubmed/27833000.

36. Kaji DA, Zech JR, Kim JS, et al. An attention based deep learning model of clinical events in the intensive care unit. *PLoS One*. 2019;14(2):e0211057. https://www.ncbi.nlm.nih.gov/pubmed/30759094.

37. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med*. 2017;89:248-255. https://www.ncbi.nlm.nih.gov/pubmed/28843829.

38. Mao Q, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*. 2018;8(1):e017833. https://www.ncbi.nlm.nih.gov/pubmed/29374661.

39. Shashikumar SP, Stanley MD, Sadiq I, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol*. 2017;50(6):739-743. https://www.ncbi.nlm.nih.gov/pubmed/28916175.

40. Taneja I, Reddy B, Damhorst G, et al. Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis. *Sci Rep*. 2017;7(1):10800. https://www.ncbi.nlm.nih.gov/pubmed/28883645.

41. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*. 2015;7(299). doi:10.1126/scitranslmed.aab3719

42. Saqib M, Sha Y, Wang MD. Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. *Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. 2018;2018:4038-4041. https://www.ncbi.nlm.nih.gov/pubmed/30441243.

43. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018;24(11):1716-1720. https://www.ncbi.nlm.nih.gov/pubmed/30349085.

44. Merouani M, Guignard B, Vincent F, et al. Norepinephrine weaning in septic shock patients by closed loop control based on fuzzy logic. *Crit Care*. 2008;12(6):R155. https://www.ncbi.nlm.nih.gov/pubmed/19068113.

45. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ open Respir Res*. 2017;4(1):e000234. https://www.ncbi.nlm.nih.gov/pubmed/29435343.

46. Rhee C, Dantes R, Epstein L, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA*. 2017;318(13):1241. doi:10.1001/jama.2017.13836

47. A computational approach to early sepsis detection. *Comput Biol Med*. 2016;74:69-73. doi:10.1016/J.COMPBIOMED.2016.05.003

48. Islam MM, Nasrin T, Walther BA, Wu C-C, Yang H-C, Li Y-C. Prediction of sepsis patients using machine learning approach: A meta-analysis. *Comput Methods Programs Biomed*. 2019;170:1-9. https://www.ncbi.nlm.nih.gov/pubmed/30712598.

49. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med*. 2018;46(4):547-553. https://www.ncbi.nlm.nih.gov/pubmed/29286945.

50. Wyk F van, Khojandi A, Davis RL, Kamaleswaran R. Physiomarkers in Real-Time Physiological Data Streams Predict Adult Sepsis Onset Earlier Than Clinical Practice. *bioRxiv*. 2018:322305. doi:10.1101/322305

51. Wacker C, Prkno A, Brunkhorst FM, Schlattmann P. Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2013;13(5):426-435. doi:10.1016/S1473-3099(12)70323-7

52. Mauri C, Principe L, Bracco S, et al. Identification by mass spectrometry and automated susceptibility testing from positive bottles: a simple, rapid, and standardized approach to reduce the turnaround time in the management of blood cultures. doi:10.1186/s12879-017-2851-5

53. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):160035. doi:10.1038/sdata.2016.35

54. Sadaka F, Ethmaneabouelmaali C, Cytron MA, Fowler K, Javaux VM, O'brien J. Predicting Mortality of Patients With Sepsis: A Comparison of APACHE II and APACHE III Scoring Systems. *Orig Artic J Clin Med Res*. 2017;9(11):907-910. doi:10.14740/jocmr3083w

55. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. *AMIA . Annu Symp proceedings AMIA Symp*. 2013;2013:1109-1115.

56. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J*. 2012;3(1):9-12. doi:10.1055/s-0031-1298595

57. Skyttberg N, Chen R, Blomqvist ; Hans, Koch S. Exploring Vital Sign Data Quality in Electronic Health Records with Focus on Emergency Care Warning Scores. 2017. doi:10.4338/ACI-2017-05-RA-0075

58. Kamaleswaran R, Akbilgic O, Hallman MA, West AN, Davis RL, Shah SH. Applying Artificial Intelligence to Identify Physiomarkers Predicting Severe Sepsis in the PICU. *Pediatr Crit Care Med*. 2018;19(10):e495-e503. https://www.ncbi.nlm.nih.gov/pubmed/30052552.

59. van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Inform*. 2019;122:55-62. https://www.ncbi.nlm.nih.gov/pubmed/30623784.

60. van Wyk F, Khojandi A, Kamaleswaran R. Improving Prediction Performance Using Hierarchical Analysis of Real-Time Data: A Sepsis Case Study. *IEEE J Biomed Heal Informatics*. 2019;23(3):978-986. doi:10.1109/JBHI.2019.2894570

# CHAPTER 5

## DIAGNOSTIC STEWARDSHIP FOR BLOOD CULTURES IN THE EMERGENCY DEPARTMENT: A MULTICENTER VALIDATION AND PROSPECTIVE EVALUATION OF A MACHINE LEARNING PREDICTION TOOL

Michiel Schinkel[1, 2*], Anneroos W. Boerman[1, 3*], Frank C. Bennis[4], Tanca C. Minderhoud[1], Mei Lie[5], Hessel Peters-Sengers[2], Frits Holleman[6], Rogier P. Schade[7], Robert de Jonge[3], W. Joost Wiersinga[2, 8], Prabath W.B. Nanayakkara[1]

[1] Section General Internal Medicine, Department of Internal Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC, location VU University Medical Center, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands.
[2] Center for Experimental and Molecular Medicine (CEMM), Amsterdam UMC, location Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
[3] Department of Clinical Chemistry, Amsterdam UMC, Vrije Universiteit Amsterdam, AGEM Research Institute, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands.
[4] Department of Computer Science, Quantitative Data Analytics Group, Department of Computer Science, Faculty of Science, VU University, De Boelelaan 1105, 1081HV, Amsterdam, The Netherlands.
[5] Department of EVA Service Center, Amsterdam UMC, location VU University Medical Center, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands and location Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
[6] Section General and Acute Internal Medicine, Department of Internal Medicine, Amsterdam UMC, location Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
[7] Department of Medical Microbiology and Infection Prevention, Amsterdam UMC, location Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
[8] Section Infectious Diseases, Department of Internal Medicine, Amsterdam UMC, location Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
* Both authors contributed equally to the manuscript.

*eBioMedicine 2022;82: 104176*
*https://doi.org/10.1016/j.ebiom.2022.104176*

# ABSTRACT

## Background

Overuse of blood cultures (BCs) in emergency departments (EDs) leads to low yields and high numbers of contaminated cultures, accompanied by increased diagnostics, antibiotic usage, prolonged hospitalization, and mortality. We aimed to simplify and validate a recently developed machine learning model to help safely withhold BC testing in low-risk patients.

## Methods

We extracted data from the electronic health records (EHR) for 44.123 unique ED visits with BC sampling in the Amsterdam UMC (locations VUMC and AMC; the Netherlands), Zaans Medical Center (ZMC; the Netherlands), and Beth Israel Deaconess Medical Center (BIDMC; United States) in periods between 2011 and 2021. We trained a machine learning model on the VUMC data to predict blood culture outcomes and validated it in the AMC, ZMC, and BIDMC with subsequent real-time prospective evaluation in the VUMC.

## Findings

The model had an Area Under the Receiver Operating Characteristics curve (AUROC) of 0.81 (95%-CI = 0.78-0.83) in the VUMC test set. The most important predictors were temperature, creatinine, and C-reactive protein. The AUROCs in the validation cohorts were 0.80 (AMC; 0.78-0.82), 0.76 (ZMC; 0.74-0.78), and 0.75 (BIDMC; 0.74-0.76). During real-time prospective evaluation in the EHR of the VUMC, it reached an AUROC of 0.76 (0.71-0.81) among 590 patients with BC draws in the ED. The prospective evaluation showed that the model can be used to safely withhold blood culture analyses in at least 30% of patients in the ED.

## Interpretation

We developed a machine learning model to predict blood culture outcomes in the ED, which retained its performance during external validation and real-time prospective evaluation. Our model can identify patients at low risk of having a positive blood culture. Using the model in practice can significantly reduce the number of blood culture analyses and thus avoid the hidden costs of false-positive culture results.



## Funding

This research project was funded by the Amsterdam Public Health – Quality of Care program and the Dutch "Doen of Laten" project (project number: 839205002).

# INTRODUCTION

Blood cultures are indispensable for diagnosing bloodstream infections (BSIs), ranking among the top seven causes of death in most European and North American countries.[1] An estimated 536.000-628.000 episodes of BSI occur annually in the United States alone, with 79.000-94.000 associated deaths.[1] Physicians tend to order blood cultures frequently due to the fear of missing such a severe but treatable condition.[2,3] In emergency departments (EDs), blood cultures are collected in many patients with suspected infections, even when the primary condition is one with a low probability of being accompanied by bacteremia, such as pneumonia or cellulitis.[2,3] Consequently, the yield of blood cultures in the ED is low.[3] The percentage of true-positive blood cultures, disregarding contamination, ranges from 1.4% to 12.2% in ED populations worldwide.[4–10] Due to these low yields, blood culture outcomes affect treatment decisions in only 0.18% to 2.8% of patients presenting to the ED with suspected infection.[4,5]



The primary goal of blood culture testing should be to maximize the identification of true BSIs. However, testing all patients with suspected infections has unwanted consequences.[11] The abundant use of blood cultures leads to unnecessarily high numbers of contaminated cultures. A substantial 40% to 55% of positive cultures can be contaminated.[5,6,8,9,12] Three decades of research on this topic

**Evidence before this study**

We performed a Pubmed title/abstract search on January 18th, 2022, using the terms "Bacteremia" OR "Bacteraemia" OR "Bloodstream Infection" AND "Machine Learning" OR "Prediction" AND "Emergency Department." The search yielded 62 papers, and we found additional articles through the references. The literature shows that various (machine learning) prediction tools for blood cultures outcomes in the emergency department (ED) have been developed. Most studies only describe the model development, while the few externally validated models are tested in at most one other center. Only the Shapiro Decision Rule seems to have made it into clinical practice.

**Added value of this study**

We have created a robust tool for predicting blood culture outcomes in the ED. The tool was validated in multiple geographical locations and various types of hospitals during the development phase and subsequently prospectively evaluated in real-time. We demonstrated a net benefit of using this tool during the real-time evaluation with a decision-curve analysis.

**Implications of all the available evidence**

The literature suggests that it is possible to predict the outcome of a blood culture that is drawn in the ED. This information can be used to substantially and safely reduce unnecessary blood culture analyses and avoid the hidden costs of false-positive culture results. We now present a robust tool that can be easily implemented in various settings, and which is already implemented in the VUMC electronic health record environment. The tool is ready to be tested in a clinical trial to formally study its impact on clinical practice.

**Research in context**

has consistently shown that contamination is associated with additional resource use (laboratory and microbiological testing), increased use of antibiotics, prolonged hospital stay, and even increased in-hospital mortality.[9,12–14]

Diagnostic stewardship interventions that provide a swift and personalized blood culture testing approach are urgently needed to reduce the overuse of blood cultures and the serious secondary effects of contamination.[15] We recently demonstrated the feasibility of using electronic health record (EHR) data in a machine learning model to detect patients at low risk of a positive blood culture, in whom blood culture analyses could safely be avoided.[10] However, this model did not lend itself well to external validation and clinical implementation due to the many features included. The current study aimed to create a simplified machine learning-based blood culture prediction tool that only uses patient characteristics, vital sign measurements, and routine laboratory results to facilitate clinical use and implementation in other hospitals. To examine the performance of this model in different care settings, we carried out a multicenter external validation in academic and teaching hospitals in various geographical locations. We also evaluated the predictions prospectively in the EHR environment and performed a decision curve analysis to establish the tool's potential net benefit to safely reduce unnecessary blood cultures.



# METHODS

## Study design, population, and data sources

We performed a retrospective multicenter study with EHR data collected from four hospitals to develop and validate a logistic regression model and a gradient-boosting decision tree model (XGBoost) for blood cultures results in the ED. The better performing XGBoost model was subsequently subjected to a prospective single-center real-time evaluation. This study adheres to the "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)".[16]

Patients were included if they were 18 years or older and underwent blood culture sampling during their ED stay. Data for developing the blood culture prediction models were extracted from the Amsterdam UMC - location VU Medical Center (VUMC) EHR system between 2016 and 2021. External validation data were extracted from the EHR systems of Amsterdam UMC – location Academic Medical Center (AMC; between 2020 and 2021) and the Zaans Medical Center (ZMC; between 2016 and 2021). We further validated the models on data from the Beth Israel Deaconess Medical Center (BIDMC; Boston, Massachusetts, United States) between 2011 and 2019, available to researchers in the online MIMIC-IV-ED database.[17] The VUMC and AMC are academic hospitals, while the ZMC and BIDMC are teaching hospitals.

For prospective real-time evaluation, the XGBoost model was further integrated into the VUMC EHR environment from EPIC (EPIC Systems Corporation, Verona, Wisconsin, United States). The model predicted blood culture results for all adults who underwent blood culture sampling in the ED. The model started predicting the probability of a positive blood culture as soon as sufficient variables were documented in the EHR (see e-Methods section on patient selection for further explanation) and updated the prediction whenever additional results came in. For the prospective evaluation in this study, we analyzed all results between October 19th, 2021, and January 25th, 2022.

Before the patients were either admitted or discharged from the ED, the final prediction was used to evaluate the model's performance. Notably, the predictions were registered in the EHR but not visible to the physicians.

## Variable selection and data preprocessing

The candidate variable selection, guided by our aim to simplify the machine learning model we created earlier, was based on the VUMC cohort.[10] We selected age, sex, vital sign measurements, and laboratory results. These variable groups were the primary predictors in the initial model and are readily available in most hospitals.[10] Based on the timestamps in the EHR, we selected only the vital signs and laboratory results that were registered in the system before the end of the ED visit. We selected laboratory tests measured in more than 50% of the patients as predictor variables. Other selection decisions were made to facilitate easy integration in different hospital systems, as discussed in the e-Methods. The most important of these selection decisions was that we only selected patient visits in which at least 20% of the vital sign data and 20% of the laboratory results were available for the prediction. Missing data was further handled using median imputation in combination with indicator variables (which indicate whether a value was measured (1) or not (0) on a patient-level), which is especially effective with data missing not at random, as is the case in our data.[18] The AMC, ZMC, and BIDMC datasets were processed similarly, and the complete preprocessing pipeline is discussed in more detail in the e-Methods of the supplementary appendix, where we also reference all the packages, modules, and libraries that were used. We cleaned the data using the R statistical software version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

## Outcome

The outcome of interest was a BSI, defined as the growth of a clinically relevant pathogen in at least one blood culture bottle collected during the ED visit. Among the cultured microorganisms, we defined contaminants based on previous literature and classified those as negative cultures.[2,6,10,19,20] e-Table 5 lists all organisms that were classified as contaminants. We also experimented with a contamination classification based on the number of bottles that grew a particular pathogen, highlighted in the e-Methods.

## Statistics

### *Model development and validation*

For the model development and validation, we used Python version 3.8.1. The VUMC cohort was randomly split into a training (80%) and test set (20%), stratified by the blood culture outcomes. Subsequently, the training data were scaled to unit variance and imputed when missing. The same standardization factor and medians of the training data were used to scale and impute the test and validation data. As with our earlier approach, we trained a logistic regression model and a gradient-boosting decision tree model, implemented through Python's XGBoost (eXtreme Gradient Boosting; XGB) library.[10] The optimal hyperparameters for both models were found through a fivefold cross-validated grid search (see e-Table 2 for further details).

We validated the models derived from the VUMC training set in the VUMC test set, AMC, ZMC, and BIDMC datasets. Therefore, according to the TRIPOD criteria, our study can be classified as both a type 2a and type 3 prediction model study.[16] The discriminatory performances were assessed using the Area Under the curve of the Receiver Operating Characteristics (AUROC) and the Area Under the Precision-Recall Curve (AUPRC). The AUPRC is more robust to class imbalances, as we see with the low incidence of positive blood culture outcomes.[21] The model calibration was assessed visually using calibration plots. Feature contributions for the logistic regression were presented using the coefficients, and those of the XGBoost model were reported using Shapley values, which correspond to the local contributions of the features for each prediction.[22]



On top of evaluating the model's performance, we analyzed the potential clinical net benefit through a decision curve analysis of the prospective real-time evaluation, as recommended by editorials in leading medical journals.[23] The net benefit decision curve analysis takes into account the relative impact of false negatives (i.e., missing a BSI) and false positives (i.e., more contaminated cultures, with associated side-effects) for a range of threshold probabilities.[23,24] A detailed description of the net benefit calculations can be found in the e-Methods.

### Ethics

The Amsterdam University Medical Centers' (UMC) local medical ethics review committee waived the review of the retrospective and prospective part of this study (IRB number: IRB00002991; case: 2020.486), as the medical research involving Human Subjects Act did not apply. De-identified data extracts were used for this study, adhering to the local privacy officer's protocol. Therefore, no informed consent needed to be obtained for the use of the data. Participant data underlying the results of this study can be shared. The data can be requested following publication of this work. The data can be shared with researchers who provide a methodologically sound proposal, which is allowed under our local privacy regulations. Proposals should be directed to the corresponding author and requestors will need to sign a data access agreement. Part of the data is available to all researchers through the MIMIC-IV-ED database (https://physionet.org/content/mimic-iv-ed/1.0)

### Role of funding source

The funding sources (Amsterdam Public Health – Quality of Care program and the "Doen of Laten" project (project number: 839205002)) had no involvement in any part of the research project and did not have any influence on the decision to submit the work for publication.

## RESULTS

### Cohort description

This multicenter development and validation study used retrospective EHR data from four hospitals (VUMC, AMC, ZMC, and BIDMC) where patients with all categories of diseases and severity presented at the ED. After selecting only adult patients who underwent blood culture sampling during their ED stay and who had over 20% of the vital signs and 20% of the laboratory variables measured, the VUMC cohort consisted of 8.027 unique visits, of whom 6.421 were randomly

allocated to the training set and 1.606 to the test set. The validation cohort sizes were 2.429 (AMC), 5.961 (ZMC), and 27.706 (BIDMC). The percentage of true-positive blood cultures ranged from 5.4% (BIDMC) to 12.3% (ZMC). The percentage of contaminated cultures, which we later classified as negative, ranged from 4.9% (BIDMC) to 10.6% (AMC). Detailed information about the predictor variables and outcomes in the different cohorts is presented in Table 1. The number of ED visits included following each step of the selection procedure is presented in e-Figure 1 and frequently found microorganisms in the different cohorts are presented in e-Figure 7.

## Training performances during the model development



Based on the AUROC and AUPRC, the XGBoost model consistently outperformed the logistic regression model. Therefore, we only present the XGBoost model performances here. A detailed description of the logistic regression model performance can be found in the supplementary appendix. The XGBoost model reached an average AUROC of 0.78 (standard deviation (SD) = 0.01) and an AUPRC of 0.34 (SD = 0.01) during the training phase, visualized in Figures 1a and 1a. The calibration plot, presented in Figure 1c, shows that the model is well-calibrated. Notably, the calibration plot comprises ten bins of equal population size. High probabilities were rare, as shown in the grey histogram of the prediction distributions in Figure 1c.

**Table 1.** Cohort descriptions of predictor variables and outcomes in the datasets used to develop and validate the XGBoost model to predict blood culture outcomes in the emergency department.

| Variable | VUMC training (n=6.421) | VUMC test (n=1.606) | AMC (n=2.429) | ZMC (n=5.961) | BIDMC (n=27.706) |
|---|---|---|---|---|---|
| Age, median, y (IQR) | 66 (52-76) | 66 (53-76) | 62 (48-73) | 71 (58-81) | 61 (49-73) |
| Sex, Female, n (%) | 3666 (43.2%) | 896 (44.2%) | 1134 (46.7%) | 2770 (46.5%) | 14075 (50.8%) |
| **Vital signs, median (IQR)** | | | | | |
| Temperature, Celsius | 37.7 (36.9-38.5) | 37.8 (36.9-38.5) | 37.0 (36.3-37.7) | 37.4 (36.6-38.3) | 36.8 (36.6-37.1) |
| Heart rate, /min | 94 (81-106) | 93 (81-105) | 90 (78-102) | 95 (83-109) | 85 (74-96) |
| Systolic blood pressure, mmHg | 124 (110-140) | 123 (110-140) | 128 (113-144) | 129 (114-145) | 125 (112-139) |
| Diastolic blood pressure, mmHg | 74 (66-83) | 74 (65-83) | 76 (67-85) | 78 (68-87) | 70 (62-78) |
| Respiratory rate, /min | 20 (16-25) | 20 (16-25) | 20 (16-24) | 20 (16-25) | 18 (16-19) |
| Saturation, % | 96 (95-98) | 96 (94-98) | 97 (95-98) | 96 (93-98) | 98 (96-99) |
| **Laboratory results, median (IQR)** | | | | | |
| C-Reactive Protein | 63 (21-141) | 58 (19-142) | 46 (12-115) | 69 (28-160) | 48 (11-113) |
| Creatinine | 85 (66-119) | 84 (65-116) | 88 (69-129) | 85 (67-114) | 88 (62-133) |
| Leukocytes | 10.4 (7.0-14.5) | 10.3 (6.9-14.5) | 9.1 (6.2-13.1) | 10.6 (7.3-14.75) | 9.3 (6.6-12.9) |
| **Outcome** | | | | | |
| Positive blood cultures, % | 11.5 | 11.5 | 11.2 | 12.3 | 5.4 |
| Contaminated cultures, % | 6.3 | 6.3 | 10.6 | 5.2 | 4.9 |

IQR = Interquartile Range; VUMC = VU Medical Center; AMC = Academic Medical Center; ZMC = Zaans Medical Center; BIDMC = Beth Israel Deaconess Medical Center.

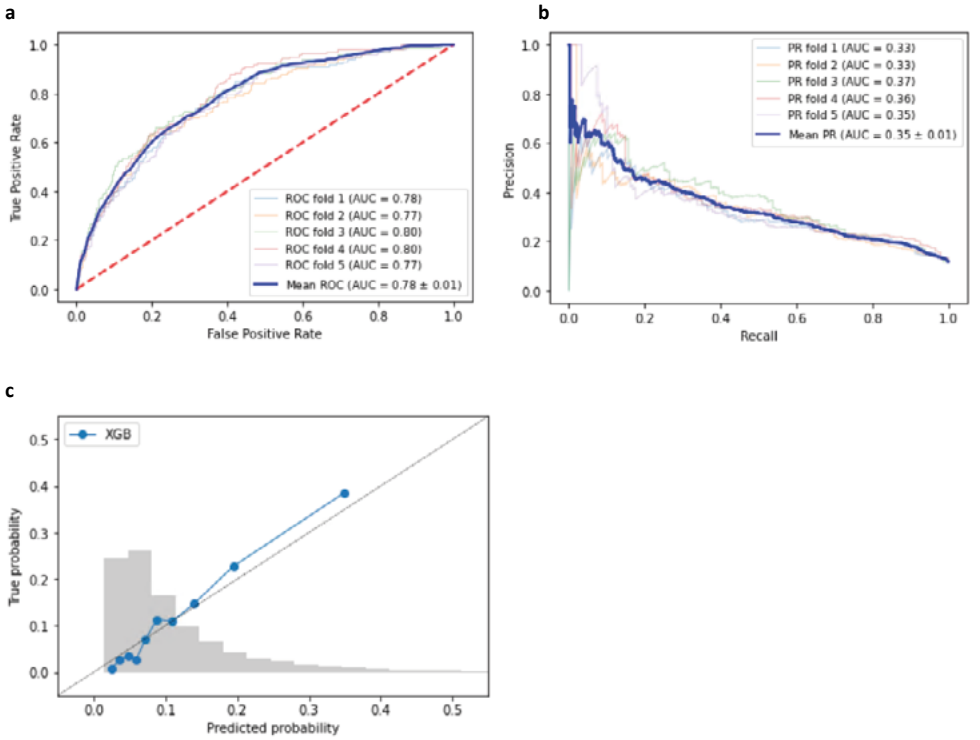



**Figure 1. Discriminatory performance and calibration of the XGBoost model for predicting the outcome of blood cultures in the emergency department in the VUMC training set: a.** the area under the receiver operating characteristics curve (AUROC). b. the area under the precision-recall curve (AUPRC). C. the calibration plot of predicted probabilities compared with actual probabilities. In grey, we further see a histogram of the distribution of the predictions in the training set in this figure.

## Features and feature importances

Based on the VUMC development cohort, we selected age, sex, six vital sign measurements, and eighteen laboratory tests as predictor variables in the model. With an additional 23 indicator variables, the model included 49 features. Details on the percentage of imputed values per feature are presented in e-Table 3. Summary statistics of the features, stratified by blood culture outcome, are presented in e-Table 4.

Figure 2a shows the twenty most important features in the XGBoost model in descending order. These features were a mixture of vital signs and laboratory results, while there was just one indicator feature among the top 20 (measurement of urea). Temperature, creatinine, and C-reactive protein were the top predictors. Figure 2b shows that low (blue) temperatures are generally associated with a negative blood culture (to the left of 0 on the x-axis), whereas high (red) temperatures are usually associated with positive blood cultures (to the right side of 0 on the x-axis).

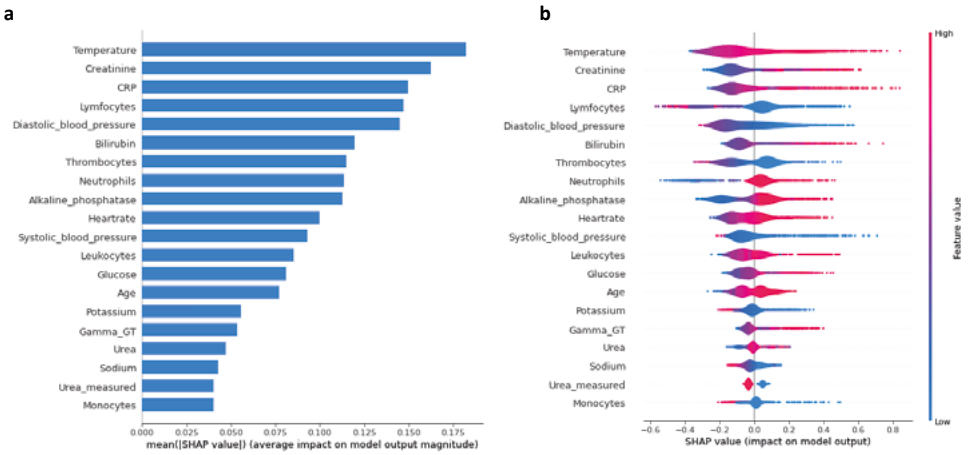



**Figure 2. Feature importances of the top 20 predictors of the XGBoost model when predicting the outcome of blood cultures in the emergency department.** According to the Shapley values, we see a. The average impact of the features on the prediction (either positive or negative). b. The local contributions of each feature for every prediction. Contributions on the left of 0 on the x-axis are associated with negative blood culture predictions, and contributions to the right of 0 on the x-axis are associated with positive blood culture predictions. The color represents the actual value of the feature at that particular prediction: blue represents a low actual value and red a high actual value.

## External validation of the prediction model

We validated the performance on the VUMC test set and external datasets from a Dutch academic medical center (AMC), a Dutch regional teaching hospital (ZMC), and a large United States-based teaching hospital (BIDMC). Figure 3a shows that the model achieves an AUROC of 0.81 (95%-CI = 0.78-0.83) within the VUMC test set and retains AUROCs of 0.80 (95%-CI = 0.78-0.82), 0.76 (95%-CI = 0.74-0.78), and 0.75 (95%-CI = 0.74-0.76) in the AMC, ZMC, and BIDMC cohorts, respectively. Figure 3b shows that the AUPRC is 0.34 (95%-CI = 0.29-0.38) in the internal test set. The AUPRC is comparable in the AMC (0.38; 95%-CI = 0.34-0.42) and ZMC (0.33; 95%-CI = 0.31-0.36), but lower in the BIDMC (0.19; 95%-CI = 0.18-0.20). Overall, the model seems to be well-calibrated in all cohorts, as seen in figure 3c.

## Prospective evaluation

Following the external validation, we integrated the XGBoost model into the EHR environment of the VUMC for a single-center real-time prospective evaluation. The model reached an AUROC of 0.76 (95%-CI = 0.71-0.81) and an AUPRC of 0.34 (95%-CI = 0.27-0.41) during the evaluation, as shown in Figure 5. In e-Figure 8, we display the pathogens found in the prospective evaluation cohort. If we had avoided blood culture draws or canceled the analysis thereof in all patients with a risk of a positive culture of less than 5%, we would have avoided 179 (30.3%) blood cultures, of which 18 gave false-positive results, and missed 5 out of 76 pathogens in the cohort.

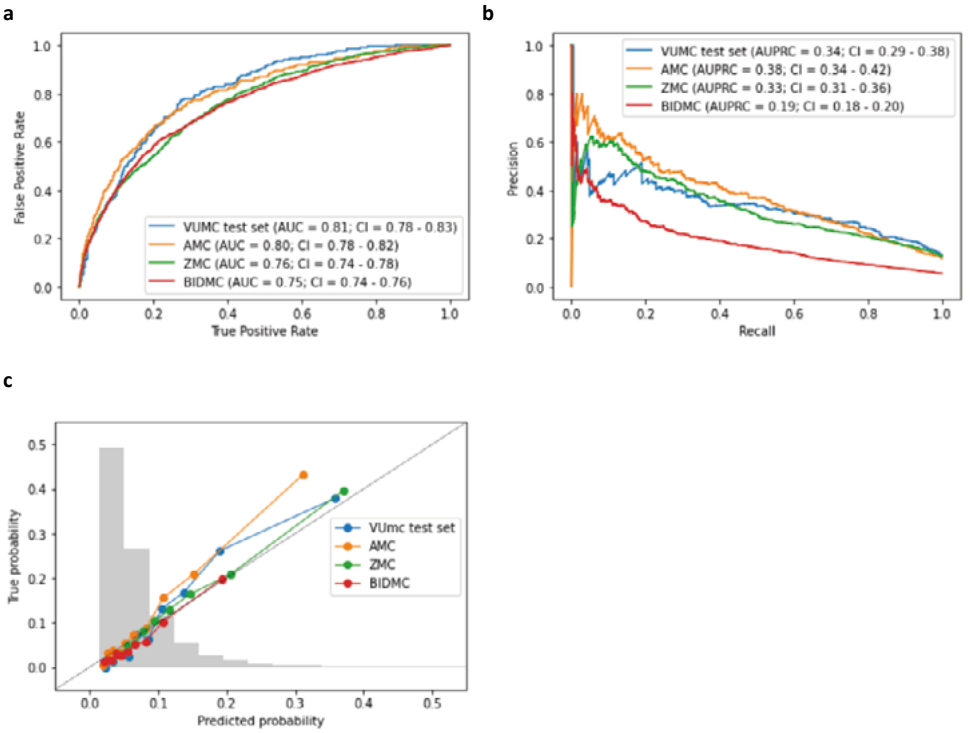



**Figure 3. Discriminatory performance and calibration of the XGBoost model for predicting the outcome of blood cultures in the emergency department during validation in the VU Medical Center (VUMC) test set, Academic Medical Center (AMC), Zaans Medical Center (ZMC), and the Beth Israel Deaconess Medical Center (BIDMC).** a. the area under the receiver operating characteristics curve (AUROC). b. the area under the precision-recall curve (AUPRC). c. the calibration plot of predicted probabilities compared with actual probabilities. In grey, we further see a histogram of the distribution of the various predictions of all four datasets combined.

## Decision curve analysis

The net benefit decision curve in Figure 5 shows that using the model to guide blood culture analyses in the ED could yield a net benefit over the current "culture all" approach across a range of threshold probabilities between 0.01 and 0.4 (40% probability of a positive culture) during the prospective real-time evaluation. According to Figure 5, the most significant benefits would be gained when thresholds between 0.1 and 0.2 would be used as cut-offs to withhold blood culture analyses. Although the net benefit at a threshold of 0.05 (5% probability of a positive culture) is much smaller, we presented the results of the prospective analysis at this cut-off since higher probabilities of missing a positive culture may not be accepted in practice. For more details on the decision curve analysis and net benefit calculations, see the e-Methods.

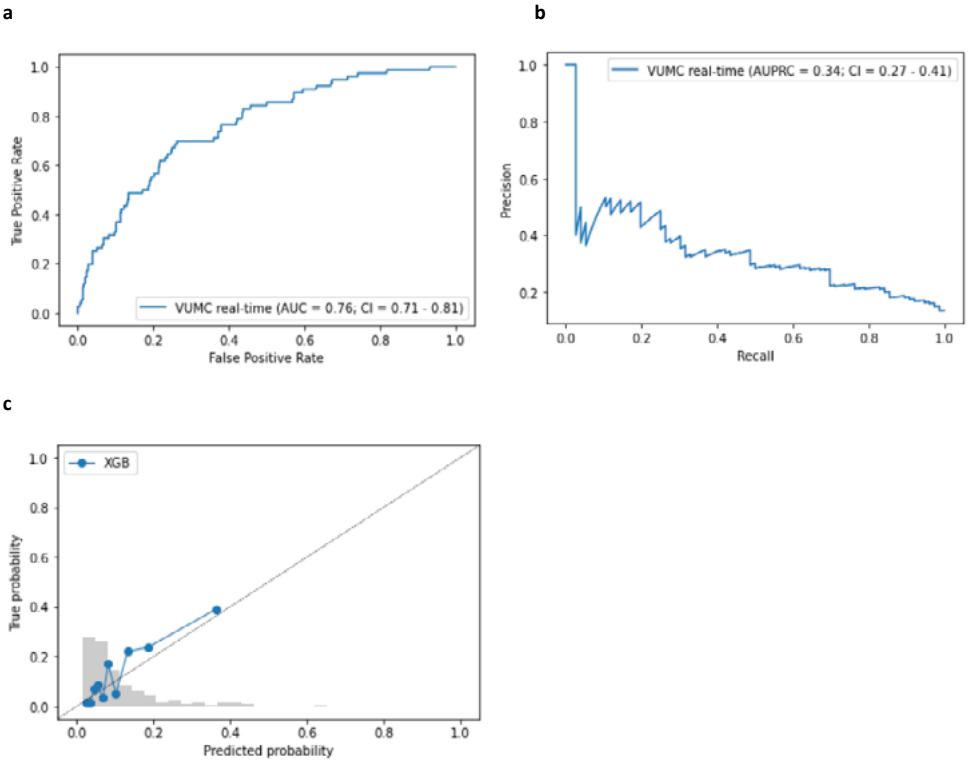



**Figure 4. Prospective evaluation of the XGBoost prediction model for blood culture outcomes in the emergency department in the VU Medical Center.** a. the area under the receiver operating characteristics curve (AUROC). b. the area under the precision-recall curve (AUPRC). c. the calibration plot of predicted probabilities compared with actual probabilities. In grey, we further see a histogram of the distribution of the predictions in the prospective evaluation.

## DISCUSSION

We created a machine learning prediction model for blood culture outcomes in the ED that performed well during internal and external validations. The XGBoost model reached an AUROC of 0.81 (95%-CI = 0.78-0.83) in the test set and up to 0.80 (95%-CI = 0.78-0.82) in external validations. Furthermore, a prospective real-time evaluation in the EHR environment of the VUMC showed that the model could retain a real-time performance with an AUROC of 0.76 (95%-CI = 0.71-0.81). A decision curve analysis showed that using the model in practice could provide a net benefit over the current approach across a large range of threshold probabilities for a positive blood culture.

Researchers have created several prediction models for blood culture outcomes in the past. Eliakim-Raz and colleagues presented fifteen such models in a 2015 systematic review.[25] Of those models, only the Shapiro decision rule seems to have been implemented in practice.[26] This striking gap between the development and implementation of prediction models has been apparent throughout the medical literature.[27,28] A review by Fleuren et al. on machine learning readiness

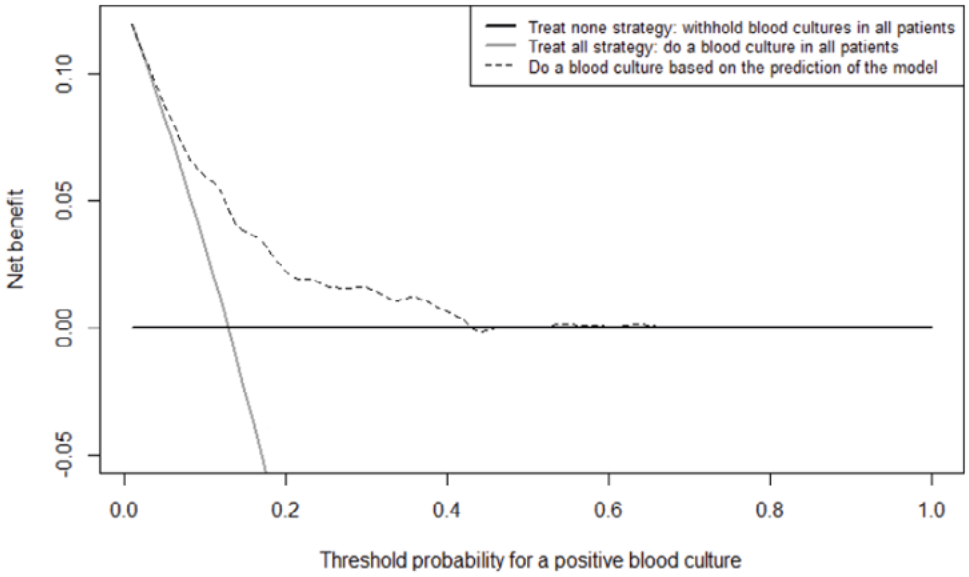



**Figure 5. A net benefit decision curve analysis of the use of the XGBoost model to decrease blood culture testing during a prospective evaluation in the VUMC.** Using the model provides a net benefit over a treat-all or treat none approach over an extensive range of potential cut-offs for converting the probability into an advice to do or withhold a blood culture.

showed that 93% of machine learning papers discuss the development of a predictive model, while just 5% externally validate the models, and only 1% do real-time testing.[28] In our study, we present a machine learning model that outperforms the current standard set by the Shapiro decision rule, and we complete stages one (problem identification) through six (real-time testing) of the machine learning readiness process.[28] Further steps will be to acquaint physicians with the prediction model in a pilot study and then perform a randomized clinical trial to establish the model's effects in practice.

Various aspects of our analyses support the validity of the predictions. Firstly, the model retained its predictive performance during external validations in different hospitals, geographical locations, and periods. Data for validation represented a mix of academic and teaching hospitals in the Netherlands and the United States. The data were captured between 2011 and 2021, including a validation set (AMC) exclusively captured during the COVID-19 pandemic. The model performance decreased slightly when we used the model in the Dutch ZMC teaching hospital or the population of Boston's BIDMC. Given the substantial differences in patient populations, outcomes, and clinical protocols, this limited performance drop is reasonable. It suggests that minor recalibrations should suffice to obtain similar performances when using the model in different hospitals.[29] The potential value of the predictions is strengthened by the comparable results we observed during the real-time prospective evaluation and is reinforced by feasible associations between the features in our model and the outcome. High temperatures, high C-reactive protein levels, and high neutrophil counts are associated with positive blood culture outcomes in our model. These variables are

all associated with BSIs and infections in general in the literature.[2,7,26] Levels of serum creatinine, bilirubin, and thrombocytes are also associated with the blood culture outcomes in our model, just as dysregulated vital sign measurements. In this case, the associations may represent critically ill patients with potential sepsis, in whom the prevalence of BSIs is known to be higher.[30]

Our machine learning tool can help reduce unnecessary blood culture analyses in the ED by identifying patients at low risk of a BSI, in whom we can safely withhold blood culture draws or cancel the analysis when the blood culture has already been sent to the lab. The consequent decrease in false-positive blood cultures may lead to lower resource use, shorter hospital stays, more appropriate use of antibiotics, and perhaps even lower in-hospital mortality.[9,12–14] Choosing which threshold probability for a positive culture is acceptable as a cut-off for doing or withholding a blood culture in practice depends on the physicians' preferences and concerns about the patient. The decision curve analysis showed that our model could provide net benefits across an extensive range of cut-offs. When using a threshold of just 5% for withholding a blood culture analysis, the model could already prevent over 30% of blood cultures, while missing a true-positive culture in 1% of cases. A clinical trial and health economic assessment are needed to fully capture the associated health- and cost gains. Choosing a higher threshold as a cut-off would help avoid even more unnecessary blood cultures, but at the cost of missing additional true positives. In the worst-case scenario, withholding blood culture sampling could lead to a missed opportunity to identify a pathogen. We showed that this scenario rarely occurs at the 5% probability threshold during the prospective evaluation. And even when it did, it could still be that the pathogens were also found through other cultures (e.g., the missed *E. coli* may also have been found through urine cultures), or that the treatment strategy would have been the same regardless of the finding of the BSI. To better understand the workflow alterations that come with using our model to avoid blood culture analyses in low-risk patients, we present two cases in Textbox 1. Strictly, all available blood culture prediction tools, including the Shapiro rule, can only validly be used in situations where the physician has already decided to do a blood culture, as they are derived from datasets of patients who underwent a blood culture draw. A valid prediction will thus need to override a clinical decision that the physician already made.

A primary limitation of this study is that we were unable to reliably examine the performance of our tool in subgroups of the population with specific comorbidities or medications. We would need data stored in free-text fields for this analysis. Arguably, the performance of our model could be worse for immunocompromised patients. This limitation warrants a detailed investigation when we test the model in a clinical trial, where we could reliably capture this information. Furthermore, we defined certain microorganisms as contaminants, while they may still represent a pathogen in specific patient groups. Examples are clinically relevant infections with coagulase-negative staphylococci (CoNS) in central line-associated BSI and prosthetic cardiac valve infections. The model must be validated separately for these patient groups in a clinical trial. A final limitation of our study is that the performance of static prediction models, including our model, could vary over time due to changes in the patient characteristics or the prevalence of positive blood cultures. When we introduce the model in practice, we expect a change in the blood culture positivity rate, as physicians may be tempted to use the model in an even broader population of patients. The performance should thus be closely monitored during implementation. We hope that future

developments will make it possible to more easily implement dynamic models that can be updated in real-time and adjust predictions based on new outcome prevalence and cohort characteristics.

In conclusion, we developed a machine learning model to predict blood culture outcomes in the ED, which retained its performance during external validation and real-time prospective evaluation. Our model can identify patients at low risk of having a positive blood culture. Using the model in practice could reduce the number of unnecessary blood cultures by at least 30% and thus avoid the hidden costs of false-positive culture results.



## CONTRIBUTORS

MS, AWB, and PWBN conceptualized this study. MS, AWB, FCB, ML, and RPS curated the data. MS, AWB, FCB, TCM, FH, RPS, RJ, WJW, and PWBN collectively investigated the data and decided on the methodology to be used. MS, AWB, FCB, and HPS conducted the formal analyses. MS and ML developed the resources. FCB, TCM, FH, RPS, RJ, WJW, and PWBN supervised various parts of the research process within their expertise. MS, AWB, TCM, FH, RJ, WJW, and PWBN acquired funding. MS, AWB, TCM, and HPS drafted the original manuscript. MS, AWB, FCB, TCM, ML, HPS, FH, RPS, DJ, WJW, and PWBN reviewed, edited, and agreed with the final version of the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests pertaining to the submitted work.

## DATA SHARING STATEMENT

Participant data underlying the results of this study can be shared after de-identification. The data can be requested following publication of this work. The data can be shared with researchers who provide a methodologically sound proposal, which is allowed under our local privacy regulations. Proposals should be directed to the corresponding author and requestors will need to sign a data access agreement. Part of the data is available to all researchers through the MIMIC-IV-ED database (https://physionet.org/content/mimic-iv-ed/1.0)

## ACKNOWLEDGMENT

We want to thank Melchior Pot (business intelligence specialist at Amsterdam UMC) for his help in implementing the prediction model in the EHR environment of the VUMC for the prospective evaluation. Furthermore, we thank Muhammad Al-Dulaimy, Bas Nieuwenhuizen, and Yvonne Bandt for providing the data from the ZMC. We are grateful for the financial contributions to this project made by the Amsterdam Public Health – Quality of Care program and the “Doen of Laten” project.

## REFERENCES


1. Goto M, Al-Hasan MN. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. Clin. Microbiol. Infect. 2013; **19**: 501–9.
2. Coburn B, Morris AM, Tomlinson G, Detsky AS. Does this adult patient with suspected bacteremia require blood cultures? *JAMA - J Am Med Assoc* 2012; **308**: 502–11.
3. Long B, Koyfman A. Best Clinical Practice: Blood Culture Utility in the Emergency Department. J. Emerg. Med. 2016; **51**: 529–39.
4. Howie N, Gerstenmaier JF, Munro PT. Do peripheral blood cultures taken in the emergency department influence clinical management? *Emerg Med J* 2007; **24**: 213–4.
5. Mountain D, Bailey PM, O'Brien D, Jelinek GA. Blood cultures ordered in the adult emergency department are rarely useful. *Eur J Emerg Med* 2006; **13**: 76–9.
6. Nannan Panday RS, Wang S, Van De Ven PM, Hekker TAM, Alam N, Nanayakkara PWB. Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. *PLoS One* 2019; **14**. DOI:10.1371/journal.pone.0214052.
7. Jessen MK, Mackenhauer J, Hvass AMW, *et al*. Prediction of bacteremia in the emergency department: An external validation of a clinical decision rule. *Eur J Emerg Med* 2016; **23**: 44–9.
8. Denny KJ, Sweeny A, Crilly J, Maloney S, Keijzers G. Is it time for a culture change? Blood culture collection in the emergency department. *Emerg Med Australas* 2018; **30**: 575–7.
9. Klucher JM, Davis K, Lakkad M, Painter JT, Dare RK. Risk factors and clinical outcomes associated with blood culture contamination. *Infect Control Hosp Epidemiol* 2021; : 1–7.
10. Boerman AW, Schinkel M, Meijerink L, *et al*. Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. *BMJ Open* 2022; **12**: 53332.
11. Garcia RA, Spitzer ED, Beaudry J, *et al*. Multidisciplinary team review of best practices for collection and handling of blood cultures to determine effective interventions for increasing the yield of true-positive bacteremias, reducing contamination, and eliminating false-positive central line-a. *Am J Infect Control* 2015; **43**: 1222–37.
12. Zwang O, Albert RK. Analysis of strategies to improve cost effectiveness of blood cultures. *J Hosp Med* 2006; **1**: 272–6.
13. Bates DW, Goldman L, Lee TH. Contaminant Blood Cultures and Resource Utilization: The True Consequences of False-Positive Results. *JAMA J Am Med Assoc* 1991; **265**: 365–9.
14. Dempsey C, Skoglund E, Muldrew KL, Garey KW. Economic health care costs of blood culture contamination: A systematic review. Am. J. Infect. Control. 2019; **47**: 963–7.
15. Curren EJ, Lutgring JD, Kabbani S, *et al*. Advancing Diagnostic Stewardship for Healthcare-Associated Infections, Antibiotic Resistance, and Sepsis. *Clin Infect Dis* 2021; published online Aug 4. DOI:10.1093/CID/CIAB672.
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med* 2015; **13**: 1.
17. Johnson, A., Bulgarelli, L., Pollard, T., Celi, L. A., Mark, R., & Horng S. MIMIC-IV-ED (version 1.0). PhysioNet. 2021. https://doi.org/10.13026/77z6-9w59.
18. Choi J, Dekkers OM, le Cessie S. A comparison of different methods to handle missing data in the context of propensity score analysis. *Eur J Epidemiol* 2019; **34**: 23–36.
19. Dargère S, Cormier H, Verdon R. Contaminants in blood cultures: importance, implications, interpretation and prevention. Clin. Microbiol. Infect. 2018; **24**: 964–9.
20. Harvey DJ, Albert S. Standardized definition of contamination and evidence-based target necessary for high-quality blood culture contamination rate audit. J. Hosp. Infect. 2013; **83**: 265–6.
21. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997; **30**: 1145–59.

22. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. http://github.com/slundberg/shap (accessed Oct 21, 2021).

23. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic Progn Res 2019 31* 2019; **3**: 1–8.

24. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models: *http://dx.doi.org.vu-nl.idm.oclc.org/101177/0272989X06295361* 2016; **26**: 565–74.

25. Eliakim-Raz N, Bates DW, Leibovici L. Predicting bacteraemia in validated models-a systematic review. Clin. Microbiol. Infect. 2015; **21**: 295–301.

26. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. Who Needs a Blood Culture? A Prospectively Derived and Validated Prediction Rule. *J Emerg Med* 2008; **35**: 255–64.

27. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med* 2021; : 1–11.

28. Fleuren LM, Thoral P, Shillan D, *et al*. Machine learning in intensive care medicine: ready for take-off? Intensive Care Med. 2020; **46**: 1486–8.

29. Davis SE, Greevy RA, Fonnesbeck C, Lasko TA, Walsh CG, Matheny ME. A nonparametric updating method to correct clinical prediction model drift. *J Am Med Informatics Assoc* 2019; **26**: 1448–57.

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

# SUPPLEMENTARY APPENDIX

## e-Methods

### *A. Patient selection*

For the VUMC, AMC, and BIDMC cohorts, we received anonymized extracts of the complete emergency department (ED) population to further filter based on our specific study. We processed all cohorts similarly and selected all adult patients for whom a blood culture was registered in the Electronic Health Records (EHR) system during the ED stay, based on the associated time stamps. The ZMC dataset was created by local business intelligence specialists and only included patients who had a blood culture taken during the ED stay. Therefore, we did not need to select these further. We excluded patients with ED stays of over 24 hours, as these may have been caused by errors in the EHR registration times or would be rare situations.



After the feature selection procedure (described below), we selected only the patients for whom at least 20% of the vital signs and 20% of the laboratory results were available, ensuring a minimum of four actual laboratory results and two vital signs as the basis for the prediction. These selections would also be made in the prospective evaluation and filters out irregularities in the data as it confirmed that a nurse had actively seen the patient and recorded vital sign data (with just one vital sign, this could have been a single automated heart rate measurement by the monitor) and also confirmed that the physician had actively ordered diagnostic tests other than the blood culture. e-Figure 1 shows an overview of the number of ED visits at each stage of the selection process.

### *B. Variable selection*

The variable selection was carried out in the VUMC cohort, as this was the development set. We selected age, sex, laboratory results, and vital sign measurements, as these variable groups were the primary drivers of the predictions in the initial model [1]. For the laboratory results and vital sign measurements, we selected only those whose outcomes were registered in the EHR system before the end of an ED visit. We averaged the results if there were multiple measurements of a variable during one visit. From the complete set of laboratory results measured in the VUMC population of patients who underwent a blood culture draw in the ED, we only used variables measured in over 50% of the population. We further disregarded the estimated Glomerular Filtration Rate (eGFR) as a predictor variable, as it could be calculated differently in different hospitals. When variables were measured in 30% to 50% of the population, we created indicator variables to denote when these were measured without using the actual values. This was the case for albumin, aspartate aminotransferase, and lactate dehydrogenase.

### *C. Variable cleaning*

As we work with EHR data, there can be erroneous measurements included in the data. Before finalizing the dataset, we excluded values that were deemed physiologically implausible. Cut-offs for these implausible values were based on the cut-offs by the VUMC clinical chemistry department, earlier work on this topic, or based on expert opinion [1]. The logical minimum and maximum per feature are presented in e-Table 1. For values in the BIDMC cohort, from the online MIMIC-IV-ED database, we had to convert some of the measurements to the SI units used in the Dutch populations.

### *D. Approach to missing data*

After finalizing the list of predictor variables, we added indicator variables for all laboratory tests and vital signs to indicate whether they were measured or missing on a patient level as we would need to impute the missing data during the modeling phase. We chose to impute missing values with the median of the training set, as median imputation combined with indicator variables is a practical approach to handle missing values and is especially effective with data missing not at random, as is the case in our data [2]. The percentage of missing values per variable per cohort is presented in e-Table 3.



As a sensitivity analysis, we also ran the XGBoost training on a dataset imputed using an iterative imputer [3]. The iterative imputer implements Multivariate Imputation by Chained Equations (MICE) but only returns a single imputation [4]. The results of this analysis are presented in e-Figure 4, which shows nearly identical results to the median imputation strategy. Since there was no difference in these outcomes, we opted to use the more straightforward and easier to explain median imputation approach.

### *E. Outcome processing and definitions*

The outcomes of the blood cultures could be textually different between the various cohorts. To aid reproducibility and robustness, we used the AMR package in the R statistical software to reclassify the blood culture outcomes of the various cohorts to standardized family names and microbe names [5]. We defined likely contaminants based on previous literature and classified those as negative cultures [6–9]. The list of microorganisms categorized as likely contaminants can be seen in e-Table 4, and e-Figure 7 visualizes the top 10 organisms found in the different cohorts.

Since classification based on the microorganism but not the clinical context could introduce a bias, we experimented with a different approach. We defined contamination based on the number of bottles with the likely contaminants as a sensitivity analysis. When likely contaminants were found in over 50% of culture bottles, they were classified as positive in this sensitivity analysis. One problem with this approach was that 57.8% of the population had only one set (of two bottles) of blood cultures taken during their visit, compared with 34.9% who had two sets of blood cultures drawn, and just 7.3% with three or more sets. In the 57.8% of cases with just one set, and thus two bottles, a likely contaminant was already classified as positive if they were present in one bottle. Therefore, the total number of cultures classified as positive increased substantially from 922 to 1121. e-Figure 5 shows that the model's performance with these outcome labels was considerably worse in the VUMC training cohort, with an AUROC of 0.75. Since this approach would also present difficulties when only an overall blood culture outcome would be shown in the EHR system, instead of per bottle, it was deemed inferior.

### *F. Model training*

The VUMC cohort was split into a training (80%) and test (20%) set, stratified by outcomes. After scaling and imputing the data through a pipeline, we trained a logistic regression and XGBoost model on the VUMC training cohort [10–14]. The optimal hyperparameters were found through a fivefold cross-validated grid search, of which further details are presented in e-Table 1 [15, 16].

### *G. Logistic regression model outcomes*

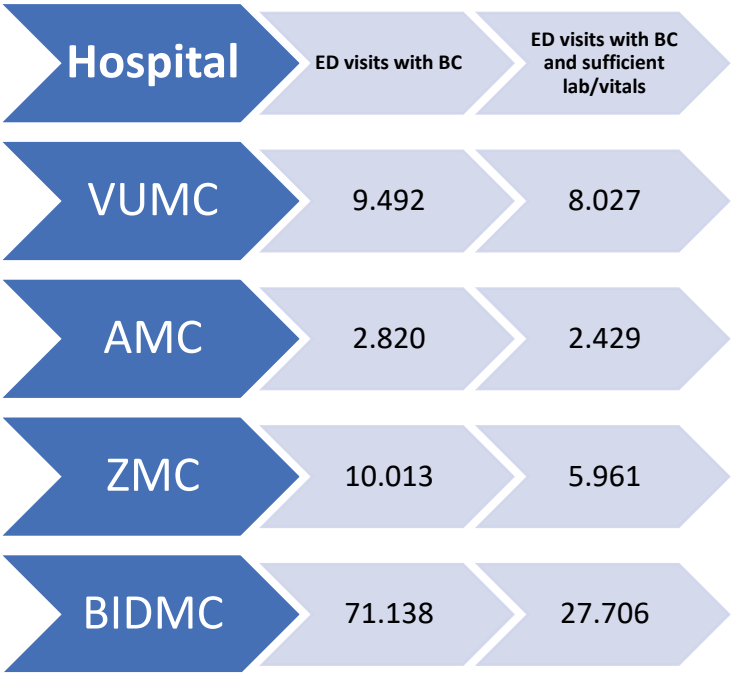
Since the XGBoost model consistently outperformed the logistic regression, we presented only the XGBoost model performance in the main paper with a prospective evaluation. Here we present the logistic regression model results. During the training phase, the logistic regression model reached an Area Under the Receiver Operating Characteristics curve (AUROC) of 0.77 and an Area Under the Precision-Recall Curve of 0.31, as visualized in e-Figure 1A and 1B. e-Figure 1C shows that the model was well-calibrated. The same results for the validation phase are shown in e-Figure 2, with AUROCs ranging between 0.73-0.79 and AUPRCs between 0.16-0.32. In Figure 3, we see the importance of all the features in the logistic regression model based on the coefficients.



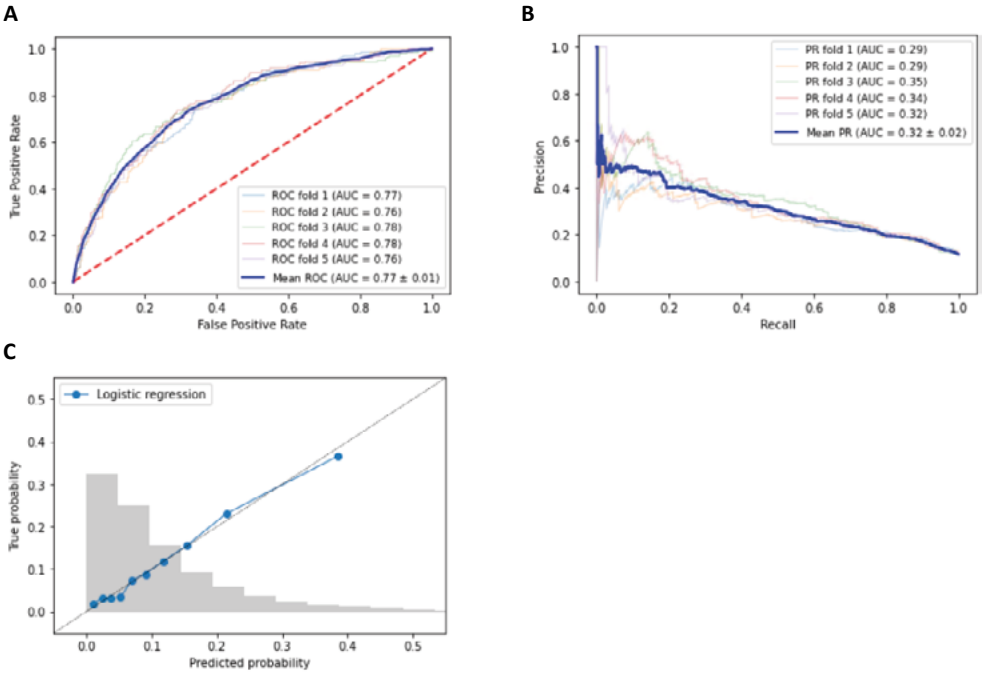
### *H. Decision curve analysis*

The benefit of using a prediction model in practice depends on the balance between the clinical benefit of doing a blood culture (i.e., not missing a positive culture, being able to narrow the antibiotic spectrum) and the adverse side effects (i.e., more contaminated cultures with associated resource use, antibiotics use, increased length of hospital stay) [17, 18]. As our model outputs a probability of a positive blood culture, we can define a threshold probability where the expected benefit of taking the blood culture equals the expected benefit of withholding the culture. However, this also depends on the likelihood of the outcomes and the judgment of these outcomes by the physician. The net benefit can be calculated as (TP— w*FP)/N, where TP is the number of true-positive decisions, FP is the number of false-positive decisions, N is the total number of patients and w is a weight equal to the odds of the cut-off given by the threshold probability [17, 18]. The numbers outputted from this net benefit calculation provide a purely theoretical evaluation that does not translate perfectly into the clinical setting. However, it gives a robust evaluation of the potential approaches. At a threshold probability for a positive culture of 5%, the model achieves a net benefit of 0.088 over withholding cultures for all patients, equivalent to detecting 8.8 true-positive blood cultures per 100 patients without increasing false positives. Furthermore, the net benefit of using the model at a probability threshold of 5% is 0.006 higher than with doing blood cultures in all patients (0.088 - 0.082 = 0.006). Using the model with this cut-off would thus result in finding six additional true-positive blood culture per 1000 patients without increasing the number of false positives, compared with the current “culture all” approach. Notably, the “culture all approach” line represents the relation between the threshold used to intervene (shown on the x-axis), and the prevalence of positive blood cultures in the studied population.

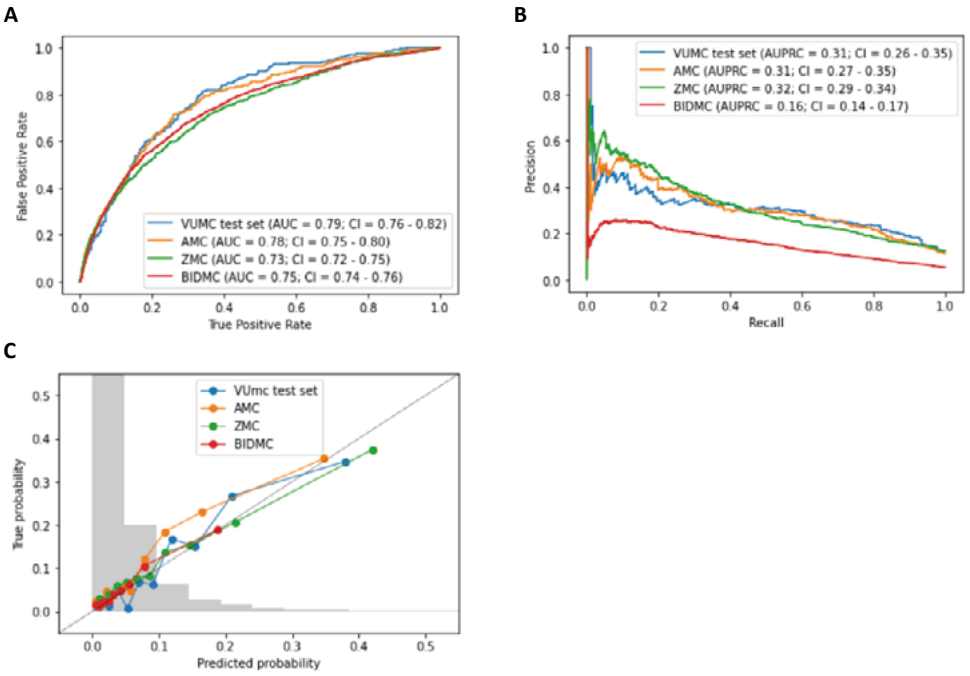
## e-Figures



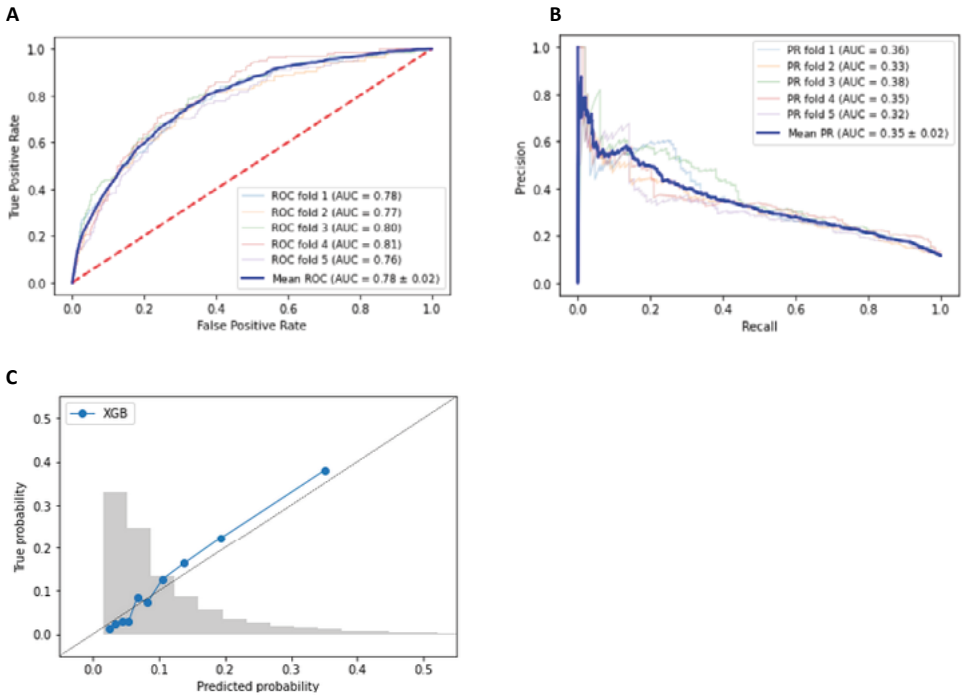

**e-figure 1. Overview of the number of emergency department visits at each stage of the data cleaning procedure in the different cohorts.** ED = Emergency Department; BC = Blood culture; VUMC = VU Medical Center; AMC = Academic Medical Center; ZMC = Zaans Medical Center; BIDMC = Beth Israel Deaconess Medical Center. Sufficient lab/vitals indicates that at least 20% of the vital sign measurements and 20% of the laboratory results were registered in the system before the end of an ED visit. The percentage of visits with sufficient data is considerable less in the BIDMC cohort, which may reflect differences in diagnostic protocols in the ED.



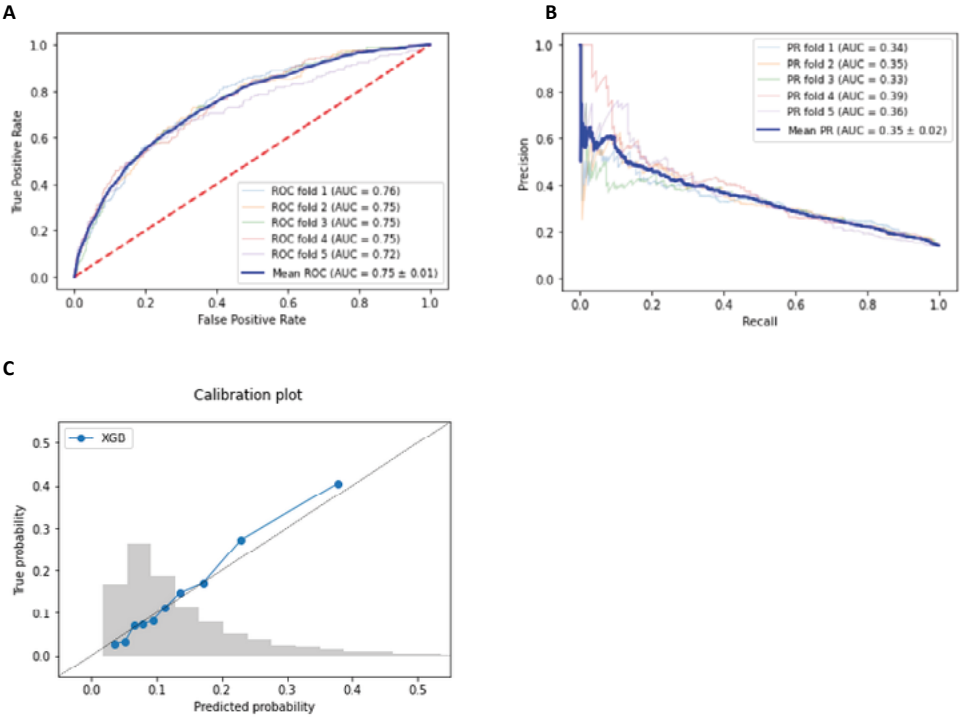

**e-Figure 2. The performance of the logistic regression model during training.** During training in the VUMC cohort, we see A. The Area Under the curve of the Receiver Operating Characteristics (AUROC). B. The Area Under the Precision-Recall Curve (AUPRC). C. The model calibration with in grey the distribution of the predictions.



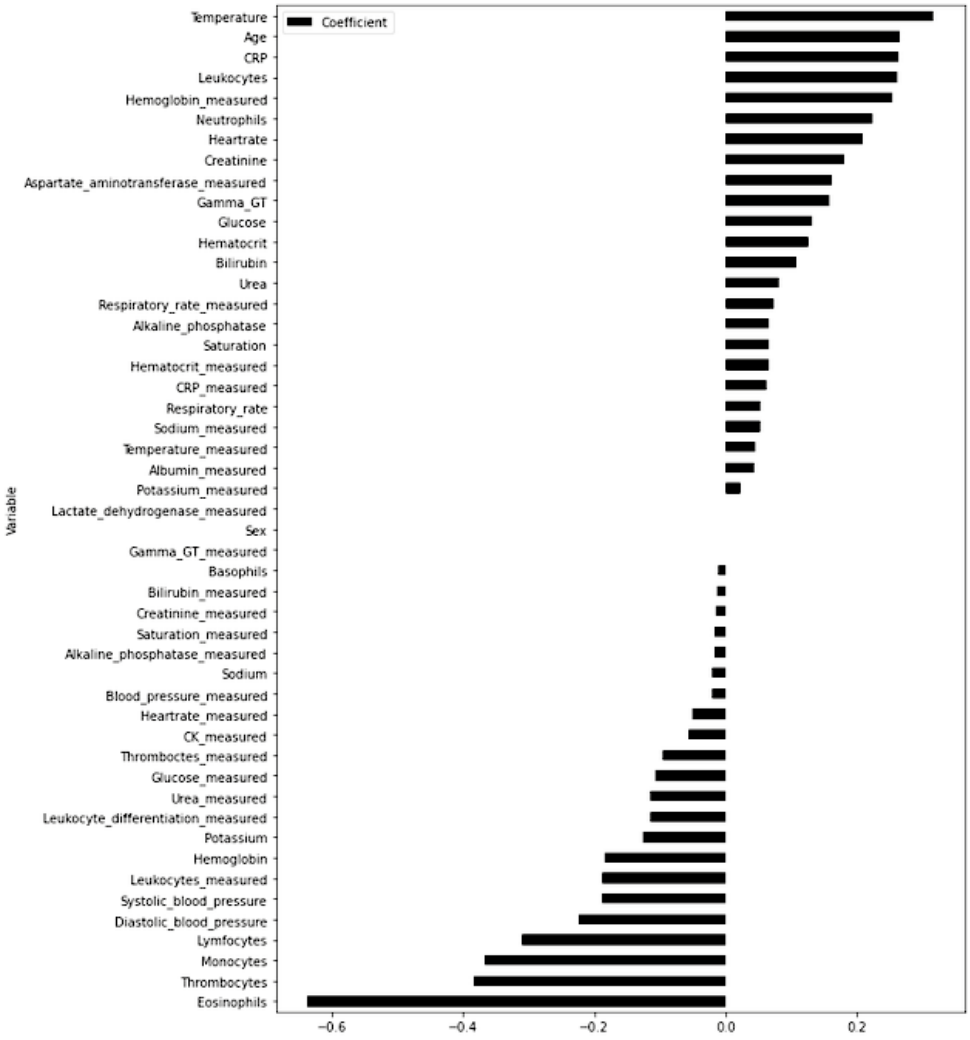

**e-Figure 3. The performance of the logistic regression model during validation.** During validation in the VUMC test set, AMC, ZMC, and BIDMC, we see A. The Area Under the curve of the Receiver Operating Characteristics (AUROC). B. The Area Under the Precision-Recall Curve (AUPRC). C. The model calibration with the distribution of the predictions in all cohorts combined in grey.



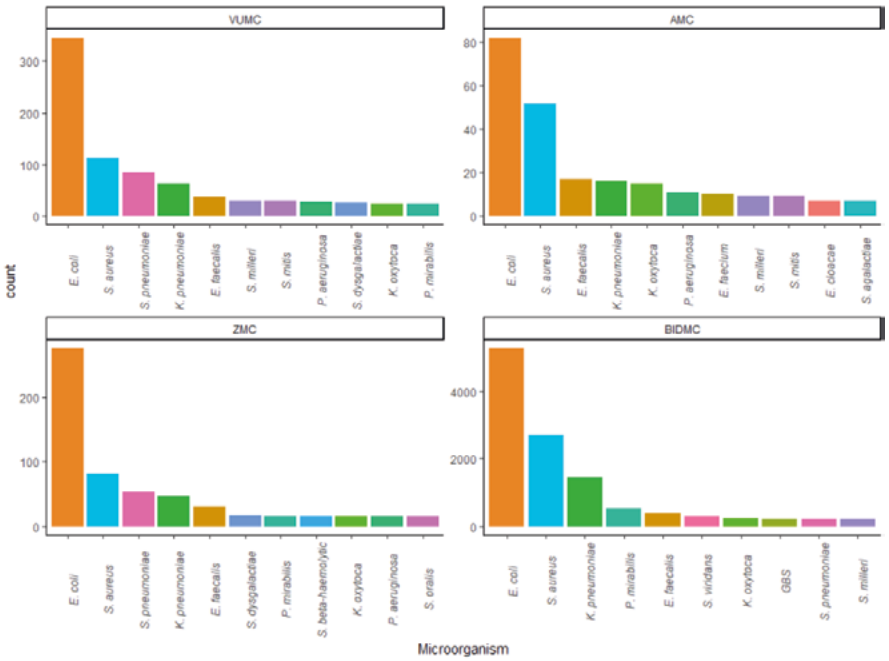

**e-Figure 4. The performance of the XGBoost model in a sensitivity analysis using iterative imputation.** Discriminatory performance and calibration of the XGBoost model for predicting the outcome of blood cultures in the emergency department. Instead of using median imputation, this analysis uses an iterative imputer. During the training phase in the VUMC training set, we see A. the Area Under the Receiver Operating Characteristics curve (AUROC). B. the Area Under the Precision-Recall Curve (AUPRC). C. the calibration plot of predicted probabilities compared with actual probabilities. In grey, we further see a histogram of the distribution of the predictions in the training set in this figure.



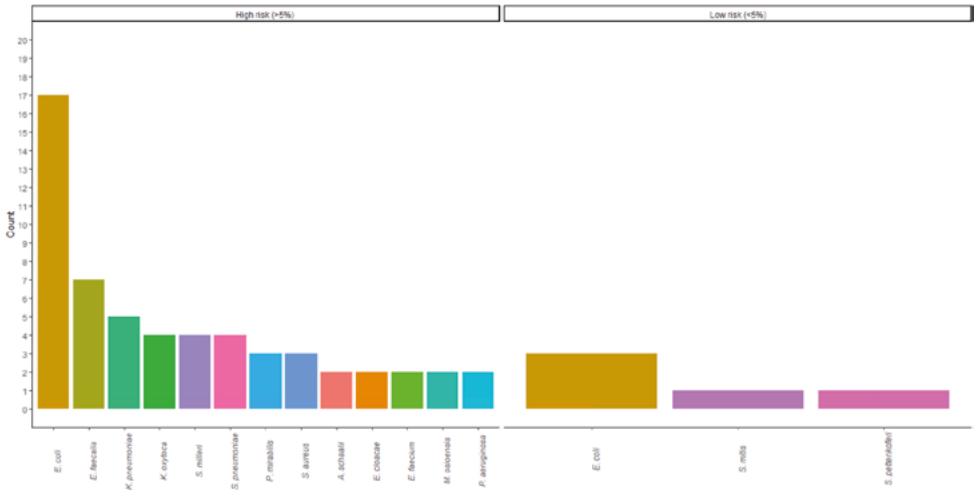

**e-Figure 5. The performance of the XGBoost model in a sensitivity analysis defining contamination based on the number of positive bottles.** Discriminatory performance and calibration of the XGBoost model for predicting the outcome of blood cultures in the emergency department. Instead of using a standard definition of contamination based on the list in e-Table 4, this analysis also considers the number of positive bottles with this likely contaminant. During the training phase in the VUMC training set, we see A. the Area Under the Receiver Operating Characteristics curve (AUROC). B. the Area Under the Precision-Recall Curve (AUPRC). C. the calibration plot of predicted probabilities compared with actual probabilities. In grey, we further see a histogram of the distribution of the predictions in the training set in this figure.




**e-Figure 6. The coefficients of the logistic regression model for all predictor variables.** This figure presents the coefficients of all the features in the logistic regression model. The features with a "_measured" suffix are the binary indicator variables (was this variable measured (1) or imputed (0). Features with a negative coefficient predict a lower chance of a positive blood culture with higher actual values of that specific variable.




**e-Figure 7. An overview of the top 10 microorganisms found in the various cohorts.** This faceted figure presents the top 10 pathogenic microorganisms found in the various cohorts. We only present pathogens and not contaminants. The number of pathogens found is higher than the number of positive tests, as some cultures contain multiple microorganisms. GBS = Group B Streptococci.




**e-Figure 8. An overview of the microorganisms found during the prospective evaluation.** The figure presents the pathogens found during the prospective evaluation and splits them based on the risk prediction by the model. The figure shows that only five pathogens were found in the low-risk group. In the high-risk group, 76 pathogens were found, of which the top 10 are presented here.

## e-Tables

**e-Table 1.** The logical minimum and maximum per variable.

| Parameter | Minimum | Maximum |
|---|---|---|
| Alkaline Phosphatase | 10 | 2500 |
| Basophils | 0.001 | 6 |
| Bilirubin | 2 | 1000 |
| Creatinine | 10 | 2500 |
| C-Reactive Protein | 0.01 | 700 |
| Eosinophils | 0 | 45 |
| Gamma GT | 3 | 6000 |
| Glucose | 1 | 120 |
| Hemoglobin | 1 | 17 |
| Hematocrit | 0.05 | 0.75 |
| Leukoctyes | 0.001 | 60 |
| Lymfocytes | 0.001 | 22 |
| Monocytes | 0.001 | 15 |
| Neutrophils | 0.001 | 60 |
| Potassium | 1 | 10 |
| Sodium | 95 | 180 |
| Thrombocytes | 0.1 | 2400 |
| Urea | 0.5 | 95 |
| Heartrate | 1 | 300 |
| Systolic blood pressure | 40 | 250 |
| Diastolic blood pressure | 40 | 250 |
| Temperature | 28 | 45 |
| Respiratory rate | 1 | 80 |
| Saturation | 15 | 100 |

**e-Table 2.** The grid of the search for optimal hyperparameters during model training.

| Hyperparameter | Model | Grid | Optimized value |
|---|---|---|---|
| Learning rate | XGBoost | [0.01, 0.05, 0.1] | 0.01 |
| Gamma | XGBoost | [1, 1.5, 2] | 1 |
| Minimum child weight | XGBoost | [1, 5, 10] | 1 |
| Maximum depth | XGBoost | [3, 5, 7] | 7 |
| Subsample | XGBoost | [0.4, 0.6, 0.8, 1] | 0.4 |
| Colsample by tree | XGBoost | [0.6, 0.8, 1] | 0.6 |
| C | Logistic regression | [0.01, 0.05, 0.1, 0.5, 1] | 0.5 |
| Solver | Logistic regression | [lbfgs, newton-cg, liblineair, saga] | saga |
| Penalty | Logistic regression | [L1, L2] | L1 |

**e-Table 3.** An overview of the percentage of imputed values in the various cohorts. The percentage of imputed values per variable per cohort. The final row shows the average rate of imputed values in the complete cohort.

| | VUMC | AMC | ZMC | BIDMC |
|---|---|---|---|---|
| **Eosinophils** | 62.64 | 44.17 | 89.41 | 55.20 |
| **Neutrophils** | 56.16 | 31.41 | 85.22 | 45.11 |
| **Basophils** | 53.72 | 30.71 | 85.93 | 51.49 |
| **Gamma_GT** | 48.30 | 38.66 | 18.44 | 99.56 |
| **Monocytes** | 43.98 | 19.35 | 84.42 | 45.41 |
| **Lymfocytes** | 43.22 | 18.69 | 82.45 | 45.46 |
| **Urea** | 40.36 | 13.42 | 16.54 | 5.15 |
| **Respiratory_rate** | 36.02 | 20.87 | 43.60 | 0.12 |
| **Bilirubin** | 30.55 | 8.77 | 20.45 | 54.48 |
| **Hematocrit** | 30.00 | 79.13 | 82.15 | 8.84 |
| **Alkaline_phosphatase** | 22.65 | 14.00 | 48.20 | 53.50 |
| **Glucose** | 16.07 | 7.53 | 28.87 | 4.78 |
| **Potassium** | 9.01 | 3.66 | 27.16 | 4.99 |
| **Sodium** | 8.07 | 1.28 | 23.74 | 4.87 |
| **Temperature** | 6.12 | 12.39 | 20.28 | 3.16 |
| **Thrombocytes** | 3.44 | 5.43 | 20.01 | 9.11 |
| **Saturation** | 3.13 | 2.31 | 2.95 | 1.49 |
| **Diastolic_blood_pressure** | 1.71 | 2.55 | 2.28 | 0.57 |
| **Systolic_blood_pressure** | 1.49 | 2.06 | 1.61 | 0.09 |
| **Creatinine** | 1.46 | 1.24 | 6.34 | 4.87 |
| **Heartrate** | 1.00 | 1.07 | 0.84 | 0.03 |
| **CRP** | 0.67 | 1.19 | 10.12 | 92.55 |
| **Leukocytes** | 0.62 | 0.66 | 2.99 | 9.17 |
| **Hemoglobin** | 0.60 | 0.99 | 3.34 | 8.68 |
| **Total percentage imputed** | 21.71 | 15.06 | 33.64 | 25.36 |

**e-Table 4.** Summary characteristics of the features stratified by blood culture outcome. A summary of all the real values of the features in the training cohort (VUmc).

| Characteristic | Culture negative (n=5683) | Culture positive (738) |
|---|---|---|
| Age (median, IQR) | 66 (52-76) | 69 (58-78) |
| Sex (%, female) | 43.2 | 40.4 |
| Alkaline phosphatase (median, IQR) | 87.5 (68.1-123.4) | 104.5 (78.8-177.9) |
| Basophils (median, IQR) | 0.03 (0.02-0.05) | 0.03 (0.02-0.05) |
| Bilirubin (median, IQR) | 8.7 (5.9-13.6) | 13.3 (7.9-22.2) |
| Creatinine (median, IQR) | 93.1 (64.8-114.6) | 103.1 (72.7-152.1) |
| C-Reactive Protein (median, IQR) | 59 (19-132) | 104 (40-213) |
| Eosinophils (median, IQR) | 0.06 (0.02-0.15) | 0.03 (0.01-0.08) |
| Gamma Glutamyltransferase (median, IQR) | 44 (24-106) | 71.5 (31.1-189.8) |
| Glucose (median, IQR) | 6.81 (5.87-8.53) | 7.5 (6.1-9.8) |
| Hemoglobin (median, IQR) | 7.7 (6.7-8.6) | 7.3 (6.4-8.4) |
| Hematocrit (median, IQR) | 0.38 (0.33-0.42) | 0.36 (0.31-0.41) |
| Leukocytes (median, IQR) | 10.2 (6.9-14.3) | 11.9 (8.0-16.6) |
| Lymphocytes (median, IQR) | 0.96 (0.57-1.50) | 0.55 (0.32-0.91) |
| Monocytes (median, IQR) | 0.71 (0.44-1.02) | 0.61 (0.29-0.99) |
| Neutrophils (median, IQR) | 7.2 (4.6-10.6) | 10.8 (7.4-14.2) |
| Potassium (median, IQR) | 4.1 (3.7-4.4) | 4.0 (3.7-4.4) |
| Sodium (median, IQR) | 137.4 (134.7-139.7) | 136.4 (133.6-139.0) |
| Thrombocytes (median, IQR) | 236 (176-315) | 209 (149-273) |
| Urea (median, IQR) | 6.4 (4.5-9.6) | 9.2 (5.9-13.9) |
| Heartrate (/min, median, IQR) | 93 (81-105) | 98.8 (86.8-111.5) |
| Systolic blood pressure (mmHg, median, IQR) | 125 (111-141) | 116 (102-133) |
| Diastolic blood pressure (mmHg, median, IQR) | 75 (66-84) | 69 (60-77) |
| Temperature (Celsius, median, IQR) | 37.7 (36.9-38.4) | 38.1 (37.3-38.9) |
| Respiratory rate (/min, median, IQR) | 20 (16-25) | 22 (18-26) |
| Saturation (%, median, IQR) | 96 (94.5-98) | 36 (94-97) |

**e-Table 5.** Microorganisms categorized as contaminants. A list of all microorganisms we categorized as contamination based on the literature [1, 6–9].

| Contaminants |
|---|
| *Staphylococci* (other than *S. aureus/S. lugdunensis/S. saprophyticus/S. pettenkoferi*) |
| *Micrococcus spp.* |
| *Propionibacterium spp.* |
| *Corynebacterium spp.* |
| *Bacillus spp.* |
| *Clostridium perfringens* |

## REFERENCES


1. Boerman AW, Schinkel M, Meijerink L, et al (2022) Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. BMJ Open 12:53332
2. Choi J, Dekkers OM, le Cessie S (2019) A comparison of different methods to handle missing data in the context of propensity score analysis. Eur J Epidemiol 34:23–36
3. 6.4. Imputation of missing values — scikit-learn 1.0.1 documentation. https://scikit-learn.org/stable/modules/impute.html. Accessed 15 Dec 2021
4. sklearn.impute.IterativeImputer — scikit-learn 1.0.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html. Accessed 15 Dec 2021
5. Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M (2021) Antimicrobial Resistance Data Analysis [R package AMR version 1.7.1]. Int J Syst Evol Microbiol 70:5607 5612
6. Coburn B, Morris AM, Tomlinson G, Detsky AS (2012) Does this adult patient with suspected bacteremia require blood cultures? JAMA - J Am Med Assoc 308:502–511
7. Nannan Panday RS, Wang S, Van De Ven PM, Hekker TAM, Alam N, Nanayakkara PWB (2019) Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. PLoS One. https://doi.org/10.1371/journal.pone.0214052
8. Dargère S, Cormier H, Verdon R (2018) Contaminants in blood cultures: importance, implications, interpretation and prevention. Clin Microbiol Infect 24:964–969
9. Harvey DJ, Albert S (2013) Standardized definition of contamination and evidence-based target necessary for high-quality blood culture contamination rate audit. J Hosp Infect 83:265–266
10. sklearn.linear_model.LogisticRegression — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html. Accessed 22 Oct 2021
11. Python Package Introduction — xgboost 1.6.0-dev documentation. https://xgboost.readthedocs.io/en/latest/python/python_intro.html. Accessed 22 Oct 2021
12. sklearn.preprocessing.StandardScaler — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Accessed 22 Oct 2021
13. sklearn.impute.SimpleImputer — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html. Accessed 22 Oct 2021
14. sklearn.pipeline.Pipeline — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html. Accessed 22 Oct 2021
15. sklearn.model_selection.StratifiedKFold — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html?highlight=stratified kfold#sklearn.model_selection.StratifiedKFold. Accessed 22 Oct 2021
16. sklearn.model_selection.GridSearchCV — scikit-learn 1.0 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. Accessed 22 Oct 2021
17. Vickers AJ, Elkin EB (2016) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models: http://dx.doi.org.vu-nl.idm.oclc.org/101177/0272989X06295361 26:565–574
18. Vickers AJ, van Calster B, Steyerberg EW (2019) A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic Progn Res 2019 31 3:1–8

# CHAPTER 6

## IMPLEMENTING ARTIFICIAL INTELLIGENCE IN CLINICAL PRACTICE: A MIXED-METHOD STUDY OF BARRIERS AND FACILITATORS

Bo Schouten[1,*], Michiel Schinkel[2,3,*], Anneroos W. Boerman[2,5], Petra van Pijkeren[2], Maureen Thodé[2], Marlou van Beneden[4], Rishi Nannan Panday[2], Robert de Jonge[5], W. Joost Wiersinga[3,6], Prabath WB Nanayakkara[2,**]

[1] Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Public and Occupational Health, Amsterdam Public Health research institute, De Boelelaan 1117, Amsterdam, The Netherlands
[2] Amsterdam UMC location Vrije Universiteit Amsterdam, Section General Internal Medicine, Department of Internal Medicine, De Boelelaan 1117, Amsterdam, the Netherlands.
[3] Amsterdam UMC location University of Amsterdam, Center for Experimental and Molecular Medicine (CEMM), Meibergdreef 9, Amsterdam, the Netherlands.
[4] Amsterdam UMC, Department of Strategy & Development, De Boelelaan 1117, Amsterdam, the Netherlands.
[5] Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Clinical Chemistry, De Boelelaan 1117, Amsterdam, the Netherlands
[6] Amsterdam UMC location University of Amsterdam, Division of Infectious Diseases, Meibergdreef 9, Amsterdam, the Netherlands.
*Both authors contributed equally to the manuscript.
**Corresponding author

*J Med Artif Intell 2022;5:12*
*https://dx.doi.org/10.21037/jmai-22-7*

# ABSTRACT

## Background

Though Artificial Intelligence (AI) in healthcare has great potential, medicine has been slow to adopt AI tools. Barriers and facilitators to clinical AI implementation among healthcare professionals (the end-users) are ill defined, nor have appropriate implementation strategies to overcome them been suggested. Therefore, we aim to study these barriers and facilitators, and find general insights that could be applicable to a wide variety of AI-tool implementations in clinical practice.

## Methods

We conducted a mixed-methods study encompassing individual interviews, a focus group, and a nationwide survey. End-users of AI in healthcare (physicians) from various medical specialties were included. We performed deductive direct content analysis, using the Consolidated Framework for Implementation Research (CFIR) for coding. CFIR constructs were entered into the Expert Recommendations for Implementing Change (ERIC) to find suitable implementation strategies. Quantitative survey data was descriptively analyzed.

## Results

We performed ten individual interviews, and one focus group with five physicians. The most prominent constructs identified during the qualitative interim analyses were incorporated in the nationwide survey, which had 106 survey respondents. We found nine CFIR constructs important to AI implementation: evidence strength, relative advantage, adaptability, trialability, structural characteristics, tension for change, compatibility, access to knowledge and information, and knowledge and beliefs about the intervention. Consequently, the ERIC tool displayed the following strategies: identify and prepare champions, conduct educational meetings, promote adaptability, develop educational materials, and distribute educational materials.

## Conclusions

The potential value of AI in healthcare is acknowledged by end-users, however, the current tension for change needs to be sparked to facilitate sustainable implementation. Strategies that should be used are: increasing the access to knowledge and information through educational meetings and –materials with committed local leaders. A trial phase for end-users to test and compare AI algorithms. Lastly, algorithms should be tailored to be adaptable to the local context and existing workflows. Applying these implementation strategies will bring us one step closer to realizing the value of AI in healthcare.



## Keywords

Machine learning; innovation; deployment; qualitative; quantitative;

# INTRODUCTION

The start of the Information Age in the mid-20th century has been a catalyst for the use of data in medicine (1). The early days were characterized by limited datasets, created to answer specific questions, but developments were accelerated with the widespread introduction of monitoring devices and Electronic Health Record (EHR) systems around the turn of the century (2). This enhanced labeled big-data, together with increased computing power and cloud storage, boosted the use of artificial intelligence (AI) in medicine (3). Today's healthcare professionals often feel overwhelmed by vast amounts of data from various sources (4). As medicine enters the Age of AI, there is great potential for algorithms to help make sense of all the data and augment clinical decision-making (3, 5). Algorithms can use data points from large numbers of patients to detect subtle patterns that healthcare professionals may overlook (6). These insights can support the clinical assessment of a patient, decrease diagnostic uncertainty, and improve the overall quality of care. However, medicine has been slow to adopt AI tools. Up until 2020, only 222 AI tools were approved by the US Food & Drug Administration (FDA) and 240 in Europe (of which 124 in both)(7).



The low number of approved medical AI tools is surprising considering the fact that over 50.000 studies of clinical AI model development were available in through MEDLINE alone as of October 2022, according to an interactive dashboard(8).There seems to be a significant gap between the development and deployment of AI in the healthcare industry (6). Recent reviews have shown that about 95% of the published studies on AI only address the development of a particular algorithm (9, 10). In comparison, only 1-2% of those studies evaluate the use of the algorithms against clinically relevant outcomes, and few are integrated in practice. It has been suggested that a key factor of

**Key findings**

- The current tension for change in healthcare is insufficient to facilitate AI implementation. Implementation strategies are needed to facilitate sustainable adoption.

**What is known and what is new?**

- A wide variety of AI algorithms have been created for the healthcare industry, but few make it into clinical practice. Barriers and facilitators in this process are ill-defined.
- This study revealed barriers and facilitators healthcare professionals may experience with AI (implementation) and suggests implementation strategies.

**What is the implication, and what should change now?**

- The introduction of AI tools in practice should systematically be supported by various implementation strategies, such as increasing knowledge and information through local leaders, using a trial phase to let users test and compare AI algorithms, and tailoring the tools to the local context and existing workflows.

**Highlight box**

the poor implementation of AI algorithms is the lack of inclusivity and engagement of the end-users and their domain knowledge during the development of these tools (11).

Barriers and facilitators to clinical AI implementation among healthcare professionals are ill defined, nor have they been linked to appropriate implementation strategies to overcome them. The Consolidated Framework for Implementation Research (CFIR) is a tool to identify these types of contextual influences and explain the strikingly low implementation rates of medical AI (12). Furthermore, the barriers identified by the CFIR can be entered into the Expert Recommendations for Implementing Change (ERIC) tool for implementation strategies, to create a well-tailored approach (13).

In the current study, we applied the CFIR framework to identify barriers and facilitators to AI implementation in the clinical practice. We aim to find general insights that could be applicable to a wide variety of AI-tool implementations, so that this information can be used to facilitate implementation of future AI tools in medical practice, and realize their potential to improve patient care. We present the following article in accordance with the COREQ reporting checklist.



## METHODS

We conducted a mixed-methods study consisting of three inclusion phases: individual interviews, a focus group discussion, and a nationwide survey. The Amsterdam University Medical Centers' (UMC) local medical ethics review committee waived the review of this study as the Medical Research involving Human Subjects Act did not apply (IRB number: IRB00002991; case: 2021.0396). All participants provided informed consent. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

### Inclusion phase I – individual interviews

The first phase of this study encompassed semi-structured individual interviews with the end-users (physicians) of potential AI algorithms in healthcare, regarding barriers and facilitators to clinical AI implementation. We designed a topic list for the interview, based on CFIR literature and expertise of the research team (physicians and a psychologist) (12). The topic list was structured as an hourglass: started broadly with questions and prompts on AI in healthcare in general. Then the topic narrowed down to a clinical case vignette about an AI blood culture tool to provide physicians with specific details, questions and prompts. This AI tool was recently developed by our research group and it predicts the outcomes of blood cultures in the emergency department, which may help avoid unnecessary testing and associated harmful effects(14). During the time of interview, focus group and survey (and to date) the blood culture tool was not implemented in clinical practice. We included this tool as clinical case vignette to provide interviewed physicians with real examples from a real project, and enhancing discussions. The topic list ended with AI in healthcare in the broad sense. The topic list can be consulted in additional file 1 (translated, English). The first few interviews were used to pilot the topic list, and we then concluded that the list was comprehensive, hence no additions or revisions were made to the initial list.

We included physicians from the emergency, internal medicine (including infectious disease specialists), and microbiology departments of our Amsterdam UMC hospitals (location VUmc

and AMC). We chose to target these specific specialties as they will be the main end-users of our developed AI algorithm, and they represent a substantial percentage of all hospital physicians (15). Physicians of all experience levels were eligible for inclusion, to gain a broad and generally applicable picture.

Physicians were recruited through the department secretariats with an invitational email. When interested in participation, a research team member provided additional information regarding the study, and scheduled a date for the interview. Interviews were conducted face-to-face at one of the two hospital locations, in a private room (either a meeting room, or a personal workspace). All interviews were conducted by two researchers: a female psychologist (BS) and a male medical doctor (MS). Both interviewers had interview experience, gained through previous qualitative research projects and their education. Interviewed physicians were not close colleagues of the interviewers, nor was there a prior (work) relationship. The duration of the interviews was approximately 45 minutes. Interviews were recorded with an audio recorder and field notes were taken during interviews.



We aimed to include until data saturation for this phase was deemed to be reached. The inclusion of the individual interviews ran between August 2021 and September 2021.

### Inclusion phase II – focus group

To complement the information that emerged from the individual interviews, the second phase of this study encompassed a focus group. The expected complementary value of the focus group was that it enables and stimulates interaction and discussion, which could therefore provide additional information. By performing both individual interviews and a focus group, we have the information on the topic that people are willing to share both 'publicly' and 'privately' (16). For this phase we used equal inclusion- criteria and processes as for the individual interviews. The topic list used in the focus group was identical to the individual interviews. The physicians included in the interviews were not allowed to participate in the focus group. The focus group discussion was led by an experienced focus group interviewer (MvB), and supported by BS and MS. There was no prior relationship between MvB and the focus group participants. Due to COVID-19 restrictions, the focus group was conducted online and audio recorded through Microsoft Teams. The focus group took place in September 2021 and took 90 minutes.

### Inclusion phase III – nationwide survey

Based on the themes identified in phase I and II, we created a quantitative survey (additional file 2). The aim of this survey was to rank the most prominent barriers and facilitators from the interviews, and identify those endorsed by a large population of potential AI end-users. To keep the survey concise, we only incorporated questions on important topics identified in the interviews. A power calculation for the number of survey participants was irrelevant since we did not plan to perform any statistical tests. However, we aimed to include at least 100 survey participants, to ensure we had a variety of medical specialists of different ages. The survey thus ran between December 2021 and February 2022. An anonymous link was distributed to practitioners across the country through hospital secretariats, medical associations, and social media.

### Data collection and privacy

Audio recordings of the interviews and focus group were transcribed. Transcripts or study results were not returned to participants for review and feedback. Audio recordings and transcripts were stored digitally at the Amsterdam UMC location VUmc. Characteristics data on the participants were stored in a separate file, and not included in the audio files nor transcripts. All data materials could only be accessed by the local study researchers.

As for the survey, no directly identifiable data were collected, and participants could stay completely anonymous.

### Statistical analysis



#### *Qualitative data analysis*

We performed deductive direct content analysis(16), using the CFIR (12) to code the interviews and focus group transcripts. Transcripts were coded independently by both BS and MS. After two independently coded transcripts, an interim consensus procedure followed, to ensure inter-coder agreement. After which both coders continued independently coding all transcripts, followed by an extensive consensus process, leading up to the final codes. Coding was performed in MaxQDA 2022 (VERBI Software, 2021).

To match the barriers to implementation of AI algorithms in healthcare to implementation strategies, we used the ERIC tool (13).

#### *Quantitative data analysis*

The survey was conducted using Phase Zero (Phase Zero Software, 2021). The cohort of participants was described using means and medians when appropriate. The answers to the survey questions were reported using counts and percentages. Some questions were not answered by all participants. The number of answers and total number of responses are presented with all results.

## RESULTS

We first conducted ten individual interviews with physicians, and one focus group with five physicians. Interim analyses showed data saturation was reached after these interviews and the focus group, as no novel information emerged. Table 1 reports demographic characteristics per interview- and focus group participant. The total sample consisted of 15 physicians, 33% of whom were female, with a median age of 40 (IQR: 34-45). The most prominent constructs identified during the qualitative interim analyses were incorporated in the nationwide survey. The demographic characteristics of the 106 survey respondents are presented in Table 2. Most respondents were aged between 31 and 40 (49%), they had a median of 9 years of experience (IQR: 3-17), and were mostly associated with Internal Medicine (77%).

Figure 1 shows a word cloud of CFIR constructs, coded based on the interview and focus group data (minimum frequency of 2). In the following section we will elaborate on the relevant CFIR constructs. Some CFIR constructs were rarely or never coded. Therefore, they were deemed

**Table 1.** Demographic characteristics of the interview and focus group participants

| Participant No. | Age, years | Sex | Specialty | Experience level |
|---|---|---|---|---|
| Interview | | | | |
| 1 | 33 | Female | Internal Medicine | Resident |
| 2 | 43 | Female | Internal Medicine | Specialist |
| 3 | 31 | Female | Microbiology | Researcher |
| 4 | 37 | Male | Emergency Medicine | Specialist |
| 5 | 60 | Male | Internal Medicine | Specialist |
| 6 | 43 | Female | Internal Medicine | Specialist |
| 7 | 33 | Male | Emergency Medicine | Specialist |
| 8 | 35 | Male | Microbiology | Resident |
| 9 | 38 | Male | Internal Medicine | Resident |
| 10 | 28 | Male | Intensive Care | Resident |
| Focus group | | | | |
| 1 | 47 | Male | Internal Medicine | Specialist |
| 2 | 54 | Male | Internal Medicine | Specialist |
| 3 | 51 | Male | Microbiology | Specialist |
| 4 | 40 | Female | Internal Medicine | Specialist |
| 5 | 43 | Male | Emergency Medicine | Specialist |


**Figure 1. Code cloud of all CFIR constructs mentioned during the interviews and focus group discussion.**
CFIR= Consolidated Framework for Implementation Research

**Table 2.** Demographic characteristics of the survey participants (n=106)

| Characteristic | Subgroup | No. (%) |
|---|---|---|
| Age, years | 18-25 | 2 (1.9%) |
| | 26-30 | 16 (15.1%) |
| | 31-35 | 26 (24.5%) |
| | 36-40 | 26 (24.5%) |
| | 41-45 | 15 (14.2%) |
| | 46-50 | 5 (4.7%) |
| | 51-55 | 6 (5.7%) |
| | 56-60 | 5 (4.7%) |
| | 61-65 | 5 (4.7%) |
| Specialty | Internal Medicine | 82 (77.4%) |
| | Microbiology | 15 (14.2%) |
| | Intensive Care | 3 (2.8%) |
| | Emergency Medicine | 3 (2.8%) |
| | Orthopedics | 1 (0.9%) |
| | Pulmonology | 1 (0.9%) |
| | Other | 1 (0.9%) |
| Experience with AI* | None | 25 (23.6%) |
| | Clinical use | 17 (16.0%) |
| | Research use | 18 (17.0%) |
| | Research development | 22 (20.8%) |
| | Personal interest | 42 (39.6%) |
| | Use in daily life | 44 (41.5%) |
| Clinical experience, years, median (IQR) | | 9 (3-17) |

AI = Artificial Intelligence; IQR = Interquartile range. *total exceeds 100% since multiple answers were possible (except when "None" was selected). Totals per question may not add up to the total of 106 participants, as some questions were not answered by all participants.*

irrelevant to the topic and are not included in the results. For constructs which were also addressed in the survey, the qualitative data were enriched by the corresponding quantitative data.

## Intervention characteristics – evidence strength

Interviewed physicians found evidence strength an important facilitator, although they did not agree on the type of evidence that would be sufficient. Some voiced that comprehensive retrospective and prospective validation of an algorithm would be sufficient, while others wouldn't settle for less than a high impact intervention study: *"conduct a rigorous study and publish in [high impact journal] that will be adopted worldwide."* –physician in interview 7, hereafter I7. All interviewed physicians agreed that an internationally published RCT indicating the AI tool X led to better outcomes for patient category Y would provide the best evidence and could facilitate AI implementation by enhancing trust: *"a randomized study would be a good result. That way you'll have to accept that AI has added value."* –I4, and: *"building trust in an AI algorithm has to be based on scientific research."* –I5.

There was no consensus regarding which study outcomes should be pursued in such a trial to best facilitate implementation. Some interviewed physicians considered cost-effectiveness and/

or process optimization valuable outcomes, whereas others would only be willing to deploy AI if evidence shows benefits on patient outcomes.

To study the potential benefits of AI that would be considered valuable by a large population of physicians, we asked the survey participants to rank the following topics according to importance: cost-effectiveness, work process optimization, and patient outcomes. Surveyed physicians most often ranked these items in the following order of importance: patient outcomes, work process optimization, and cost-effectiveness (56/105; 53.3%). Some found costs to be more important than work processes, ranking them: patient outcomes, cost-effectiveness, and work process optimization (25/105; 23.8%). Among those who selected a ranking in which patient outcomes were not the most important potential benefit, the most common selection was: work process optimization, patient outcomes, and cost-effectiveness (17/105; 16.2%).



## Intervention characteristics – relative advantage

To evaluate the potential benefits of AI , many interviewed physicians compared AI to their current practice. Often, AI was compared to the brain of a medical specialist, and put forward as both an influential facilitator and barrier. Some interviewed physicians argued that healthcare professionals base their decision making on certain aspects which are hard to capture in algorithms: *"the computer has not seen that patient, and I have."* –I1, and: *"The gut feeling [that physicians have], (...) is intangible and not easily measurable. You cannot put that in an algorithm."* –I2. They see this as barrier for implementing AI, as they do not believe AI has relative advantage over experienced healthcare professionals. However, for the less experienced physicians, AI was thought to be beneficial: *"old-school physicians are like encyclopedias, they see a patient and instantly know what is wrong. (..) Less experienced physicians do not have the benefit of this pattern recognition due to less experience. For them, AI can be useful."* –focus group.

Relative advantage can also be a facilitator, as some interviewed physicians voiced that the ability to analyze complex data, endless opportunities for combining data, and the speed that AI is capable of, could never be reached by the human brain. For example: "*[AI] establishes links that we cannot establish. The computer associates observations, and is less dependent on cause-effect. So when we as physicians cannot consider a logical solution because of this dependency, the model can actually draw logical conclusions."* –focus group, and: *"we figure things out after three hours, while AI could figure it out within the first hour."* –I4.

## Intervention characteristics – adaptability

The adaptability of an AI algorithm has two important aspects for the interviewed physicians. An algorithm has to be adaptable to their patient population (regarding predictive performance), and be easy to integrate with existing workflows.

Currently, adaptability is a barrier as many interviewed physicians tend to believe an algorithm would not be applicable to their patients, even when it is validated in their specific patient population. A few examples: *"I would be stubborn, (...) and think: this is not applicable to my patient."* –focus group, *"you would worry about this, even though you know [the algorithm] has*

*really been researched in this population."* – focus group, and: *"I need to know how my patients fits into the picture."* –I1 .

### Intervention characteristics – trialability

An important facilitator for sustainable implementation of AI is trialability. All interviewed physicians argued that having a trial-and-error phase to get used to an algorithm, and compare the outcomes of an algorithm to their own clinical decision making, would enhance trust. An example from I10: *"Look, everybody is talking about AUC's and how well [their algorithm] performs etc. But, in the end of the day, that does not show me what it adds in clinical practice. It needs to be implemented, it needs to be tangible. You need to see for yourself what it adds. (...) You need time to learn to trust the algorithm."*. Another example is: *"I want to see an algorithm work in practice. To be able to work with it, and know exactly how it operates. I need to build trust. (...) If we can try an algorithm for a while, we can quickly see the benefits "* –I4.

Similar results were found during the survey. When asked whether the surveyed physicians would trust an algorithm more if they would first be able to use it next to their own judgement, 76/106 (72.4%) agreed or strongly agreed, while just 18/106 (17.0%) disagreed or strongly disagreed.

### Inner setting – structural characteristics

Structural characteristics (e.g. the social architecture or maturity of an organization) also play a role during the implementation of AI algorithms in healthcare. Although all interviewed physicians believed AI has some value in healthcare, they were not unanimously convinced that the traditional character of hospitals in general is well suited to its implementation in the short term: *"hospital care is highly conservative. We as physicians are highly conservative. So, I don't know whether [implementation of AI] will go fast to be honest."* –I9.

### Inner setting – tension for change

Tension for change, either positive or negative, was one of the most influential constructs on implementation of AI algorithms in healthcare. Some interviewed physicians did not feel the need for AI algorithms, because they rather continue working in the status quo. They explain that they learned certain ways of diagnosing/treating patients during medical education, which has become a behavioural habit that is hard to deviate from: *"it is habit. We are used to doing things a certain way. It is hard to learn something new. It has to do with trust also. Because you trust the things you already do, but if you need to implement something new you need time in order to trust such an algorithm."* –I1. Another example was voiced by the physician in I10 about the blood culture algorithm: *"I could definitely accept that prediction, because 1% is a low chance so then we won't perform the blood culture because that would be unnecessary care. But, it is so deeply rooted in our workflow, so actually I would still just perform the blood culture."*.

However, many interviewed physicians did see potential value of using in AI algorithms in healthcare: *"there sure is room for improvement [in ED diagnostics], AI could play a very important role in this."* –I4, and: *"modern physicians need to be open to the idea that computers might*

*perform better than they do (...). We only have experience of a couple of years, while the computer could have experience of thousands of human years."* –I5.

For the interviewed physicians, the potential benefits of AI in healthcare lie within process optimization, time-efficiency, cost-efficiency, and enhanced patient-centeredness, patient safety, and quality of care, as voiced by I7: *"for example at the ED, but also at multiple wards. The shortage in nurses is a substantial problem, AI could help with this. But also with regard to timeliness of processes, efficiency, and maybe costs. For example, if you would perform less blood cultures, that would help. That is of course a small example, but you could broaden the scope. I think [AI] could have value in many different areas."*. Another example is: *"you could see more patients: perform the same work with less staff. That way you can spend more time on each patient, that is very important. Patients currently get way too little time with physicians. Especially on the wards, they lie in bed for 24 hours and the physician comes to see them for 5 minutes. That is too little. While the patient and their families have many questions, and need for conversation. If we would gain more time for these things, that would be golden."* -I9.

Since the qualitative results showed that interviewed physicians can feel tension for change, but that this could be both a barrier and a facilitator, we asked the survey participants specifically how they felt about our blood culture prediction tool. In 81/104 (77.9%) answers, the surveyed physicians agreed or strongly agreed that the use of our AI algorithm for blood culture indication could add value, while just 7/104 (6.7%) disagreed or strongly disagreed.

## Inner setting – compatibility

Another influential construct is compatibility. Interviewed physicians found it important that AI algorithms are implemented to support their clinical decision making, while they keep their autonomy (to overrule an algorithm) and the responsibility for their patients. In other words, AI algorithms need to be compatible to existing workflows and decision-making and enhance/support this, as opposed to taking them over completely. For example: *"in the end the physician or nurse decides what they adopt from the algorithm. But keeping that autonomy is very important."* –I10, *"I think [AI] can be of value and support us, but we shouldn't just blindly trust it."*-I1, and: *"[AI] should become some sort of advice, and not an obligation. Because, well, advices are there to sometimes not be followed."* –I7.

All interviewed physicians believed that the responsibility for patient outcomes is always that of the physician, regardless of whether they followed or deviated from the algorithm: *"the physician has the final responsibility. For example when it comes to a medical disciplinary court, the computer is not the one to get the reprimand."* –I5.

## Inner setting – access to knowledge and information

To enable trust in AI algorithms and facilitate proper implementation, the interviewed physicians felt they need access to knowledge and information about the algorithm. More specifically, they voiced the need to be informed regarding: the overall evidence, the data used (which patient populations and variables), the validation, how it operates, and how this translates to their clinical

practice. This is particularly evident for highly complex algorithms. Examples are: *"I would like to receive some sort of package: what is the rationale, what are the studies, which datasets is it developed and trained in, what is the aim."* –I6, and: *"of course AI can sometimes be challenging to explain. So you have to take end-users by the hand. I think it is most important that we understand how the model is functioning in practice."* –I3. Another example is periodic feedback information to end-users: *"to show results and effects of the AI based decisions periodically. This way we can create trust in the algorithm."* –I3. In addition, some interviewed physicians felt that they would sooner use decision support provided by algorithms when they understand how the predictions are made. An example is: *"But if I would understand why it would predict a certain outcome, then I would be more inclined to consider whether I would or would not use it."* – I1.



In the survey, we followed-up with this frequently mentioned construct and asked how surveyed physicians would like to be informed about future AI algorithms. There was a clear preference for information integrated in the existing workflow (41/106; 38.7%) (e.g. in the EHR system where the algorithms is implemented), or frequent reminders and presentations during handover moments and teaching sessions (41/106; 38.7%)). Documentation stored in separate systems, or training periods prior to implementation were less favoured with 11/106 (10.4%) positive answers each. When asked specifically how the participants would prefer AI algorithm outputs to be presented, there was no clear preference for absolute risk percentages (31/106; 29.3%), binary suggestions to take or not take a certain action (37/106; 34.9%), or risk categories (38/106; 35.9%).

### Characteristics of the individual – knowledge and beliefs about the innovation

Knowledge and beliefs about AI algorithms in healthcare was also one of the most influential constructs. There was a wide variation of attitudes, values, familiarity with facts, etc. related to AI algorithms both between and within individual interviewed physicians. Even though there was this wide variation, many interviewed physicians voiced to have little to no prior experience with AI, especially not in their clinical work: *"Maybe there are algorithms that play a role in my life. But I don't use them myself. No, not at all."* –I9.

During the nationwide survey, 25 out of 106 (23.6%) surveyed physicians shared the belief that they had never come into contact with an AI algorithm, either in their work or outside.

### Worst case scenarios

To identify out of the box barriers, additional to the CFIR constructs, we asked interviewed physicians about potential worst case scenarios in the clinical use of AI. Interviewed physicians mostly worry about adverse outcomes for patients, i.e. delayed or wrong diagnosis, suboptimal treatment, inappropriate discharge, or even death. The physician in interview 4 voiced: *"harm to the patient. Like missing an important diagnosis, that has major negative outcomes for the patient. This could turn into a complication or even adverse event."*.

Other worst case scenarios were regarding the professional stature of physicians. This includes that deployment of AI could lead to losing their job altogether, lose the enjoyable aspects of the job, or to become a 'lazy' physician. For example: *"if all you have to do is follow the [AI] model you don't*

*have to go to medical school. Then you will just sit behind your desk and approve everything [the AI model predicts]. That would be completely worthless. At least for the physician. Although maybe it would be better for patient care."* (I10).

The survey results do not fully match the findings from the qualitative part of the study in this instance. When we asked surveyed physicians whether they were worried that AI would take over the enjoyable and interesting parts of job, 74/105 (70.5%) disagreed or strongly disagreed, while just 14/105 (13.3%) agreed.

### Implementation strategies – ERIC tool

We used the CFIR to identify potential barriers to implementation of AI algorithms in healthcare, and then linked these barriers to implementation strategies using the ERIC tool. We included all nine CFIR constructs that are described above, leading to a top 3 of the following implementation strategies: *identify and prepare champions* (cumulative percentage: 280%); *conduct educational meetings* (258%); *promote adaptability* (235%). Two other important strategies are: *develop educational materials* (153%) and *distribute educational materials* (149%), due to their high individual endorsement percentages. For the total output of ERIC strategies, see additional file 3.



## DISCUSSION

This study identified barriers and facilitators to AI implementation in clinical practice. Through individual interviews and a focus group with end-users (physicians), we found nine CFIR constructs important to AI implementation: evidence strength, relative advantage, adaptability, trialability, structural characteristics, tension for change, compatibility, access to knowledge and information, and knowledge and beliefs about the intervention (12). When linking these constructs to implementation strategies using the ERIC tool, we found that the following strategies should be used for AI implementation: identify and prepare champions, conduct educational meetings, promote adaptability, develop educational materials, and distribute educational materials (13).

AI has the potential to change medicine through its ability to augment clinical decision-making by detecting subtle patterns in vast amounts of patient data, and do so tirelessly for 24 hours a day. To reach this potential, physicians need to see the *relative advantage* of integrating AI in their current practice, and feel a *tension for change*. In general, physicians acknowledge the potential value of AI in healthcare, which is a facilitator for implementation. However, physicians in our study expressed that current behavioural habits and standard practices are hard to deviate from. To create new norms and behaviour, we need to go beyond sole awareness creation (17). As highlighted by the ERIC tool, local champions are the key to success in this process. One needs committed local leaders to inspire and actively remind others to use a specific AI tool. Lasting change and sustainable implementation of AI can then be achieved through several key CFIR constructs. Firstly, *evidence strength* is important, because physicians view a peer-reviewed and internationally published article as a facilitator to AI implementation. As described in the introduction section, there are many published papers on AI algorithms, but these are neither implemented nor deployed in clinical practice (10). This suggests that an international publication in itself is not

sufficient for sustainable implementation. In addition, *access to knowledge and information* about the algorithm is essential. This was one of the most prominent CFIR constructs in our study. *Access to knowledge and information* about an AI algorithm can be realized through the following ERIC implementation strategies: *conduct educational meetings*, *develop educational materials*, and *distribute educational materials*, which can be expedited by local champions. Hence, to provide physicians with digestible information regarding an AI algorithm, it is important to develop a toolkit with manuals and other supporting material, to distribute these toolkits, and explain and educate further in meetings (13). Our data shows that physicians favour the following information in their toolkit: overall evidence, the data used, the validation, how it operates, and how this translates into clinical benefits. Both qualitative and survey data display that physicians prefer their source of knowledge and information integrated in the existing workflow, e.g. EHR system. This is in line with the CFIR construct *adaptability*, and ERIC strategy *promote adaptability* (18). Moreover, integrating AI in the EHR will promote sustainable implementation (19, 20). Besides adapting AI to existing workflows, it is necessary to provide physicians with information regarding how well an AI algorithm is adapted to their patient population. We found that physicians tend to expect that a certain algorithm does not apply to 'their' patient, even when the algorithm has been validated in similar patients. This could be due to frequency bias in physicians, leading to the belief that the frequency of patients that will fall within the small margin of error of the algorithm (and therefore lead to a wrong prediction) is much higher than it actually is. Moreover, even though evidence-based medicine is considered the gold standard in clinical reasoning, the review by Nicolini et al. show the importance of local -context and knowledge for physicians when making clinical decisions, and how this is usually valued more than evidence from research (21). This could be overcome by *access to knowledge and information*, and *trialability* (22). Trust is another important dynamic in the interaction between AI and end-users, which has been well-studied in the literature. Factors such as explainability, transparency, and interpretability seem to be key to facilitate adoption (23). Our study further adds that physicians need a trial-and-error phase in implementation to experience these factors themselves, which has been described before (22). Besides increased trust, this will allow physicians to gain expertise with the algorithm, experience clinical benefits for patients, and will help further refinement and adaptation (24). Trialability also fits within the cycle of 'plan-do-study-act', which is a tool widely used in healthcare for quality improvement (25). The broad *range of knowledge and beliefs* regarding AI underpins the importance of *trialability* and *access to knowledge and information*. Lastly, *compatibility* emerged as a primary construct which may form a barrier to AI implementation. When end-users view an intervention as threat to their autonomy, it is less likely that implementation will be successful (12). In our study, physicians feel strongly about retaining their autonomy. They argue that they should always have the final responsibility over the patients, even when AI algorithms influence their decision making. Therefore, it is important for physicians to be able to deliberately deviate from the AI recommendations, like they can do with general clinical guidelines.

The results of this study should be interpreted in the light some limitations. Firstly, the interviewees and participants of the survey were mostly physicians from the Internal Medicine, Emergency Medicine, and Microbiology departments. However, we still feel these results are

generalizable to a broader group of physicians, since most questions were not specialty dependent, and some of the themes we found have been described in other cohorts before (22). Still, it would be helpful to tailor any implementation of an AI tool to the local context and end-users, for which additional surveys and interviews in those settings are needed to confirm the generalizability of our results. Secondly, it could be possible that the survey was subject to self-selection bias, i.e. the physicians who chose to respond to the survey might have differed from the group of physicians that chose not to respond. Lastly, in our survey we included n=106 participants. We did not perform a-priori power calculation, as it was not feasible to make assumptions about effect sizes due to the novelty of the studied subject. It is therefore challenging to make a statement regarding the representativeness of our sample size. However, we do believe that this sample size is sufficient to ensure a range of variety in the participants.



## CONCLUSION

The healthcare industry has been slow to adopt AI algorithms. We identified several widely endorsed constructs important to AI in healthcare and linked them to appropriate implementation strategies. Though the potential value of AI in healthcare is acknowledged by end-users (physicians), the current tension for change is insufficient to facilitate implementation and adoption. The tension for change can be sparked by conducting educational meetings, and developing and distributing educational materials to increase access to knowledge and information. Committed local leaders are indispensable to expedite this process. Moreover, a trial phase in which physicians can test the AI algorithms and compare them to their own judgement, may further support implementation. Finally, AI developers should try and tailor their algorithms to be both adaptable and compatible with the values and existing workflows of the users. As physicians have the final responsibility for the patient, they should be able to overrule any decision of the algorithm and keep their autonomy. Applying these appropriate implementation strategies will bring us one step closer to realizing the value of AI in healthcare.

## ACKNOWLEDGEMENT

We would like to thank all the healthcare professionals who participated in this study for their time and enthusiasm.

## FUNDING

This work was supported through the 2021 innovation grant of Amsterdam Public Heath, Quality of Care program: Quality of Care Research Program (amsterdamumc.org).

## FOOTNOTE

### Reporting checklist

The authors state that the methods of this study were carried out in accordance with relevant guidelines and regulations. The authors have completed the COREQ checklist for qualitative research (additional file 4).

## Data Sharing Statement

Data (in Dutch) are available upon reasonable request after approval of the corresponding author.

## Conflicts of Interest

All authors have completed the ICMJE uniform disclosure form. The authors have no conflicts of interest to declare.

## Ethical statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy of integrity of any part of the work are appropriately investigated and resolved. All participants signed written informed consent. The Amsterdam University Medical Centers' (UMC) local medical ethics review committee waived the review of this study as the Medical Research involving Human Subjects Act did not apply (IRB number: IRB00002991; case: 2021.0396). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

## REFERENCES


1. Wartman SA, Combs CD. Medical Education Must Move From the Information Age to the Age of Artificial Intelligence. Acad Med. 2018;93(8):1107-9.
2. Brown N. Healthcare Data Growth - An Exponential Problem 2015 [Available from: https://www.nextech.com/blog/healthcare-data-growth-an-exponential-problem.
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
4. Paranjape K, Schinkel M, Nanayakkara P. Short Keynote Paper: Mainstreaming Personalized Healthcare-Transforming Healthcare Through New Era of Artificial Intelligence. IEEE J Biomed Health Inform. 2020;24(7):1860-3.
5. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ. 2019;5(2):e16048.
6. Schinkel M, Paranjape K, Nannan Panday RS, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med. 2019;115:103488.
7. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health. 2021;3(3):e195-e203.
8. Zhang J, Whebell S, Gallifant J, Budhdeo S, Mattie H, Lertvittayakumjorn P, et al. An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. Lancet Digit Health. 2022;4(4):e212-e3.
9. Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PWG. Machine learning in intensive care medicine: ready for take-off? Intensive Care Med. 2020;46(7):1486-8.
10. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 2021;47(7):750-60.
11. Verma AA, Murray J, Greiner R, Cohen JP, Shojania KG, Ghassemi M, et al. Implementing machine learning in medicine. Cmaj. 2021;193(34):E1351-e7.
12. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50.
13. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implementation Science. 2015;10(1):21.
14. Schinkel M, Boerman AW, Bennis FC, Minderhoud TC, Lie M, Peters-Sengers H, et al. Diagnostic stewardship for blood cultures in the emergency department: A multicenter validation and prospective evaluation of a machine learning prediction tool. EBioMedicine. 2022;82:104176.
15. Boerman AW, Schinkel M, Meijerink L, van den Ende ES, Pladet LC, Scholtemeijer MG, et al. Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. BMJ Open. 2022;12(1):e053332.
16. Green J TN. Qualitative Methods for Health Research: SAGE Publications Inc; 2018.
17. Gawande A. Slow Ideas The New Yorker 2013.
18. Mendel P, Meredith LS, Schoenbaum M, Sherbourne CD, Wells KB. Interventions in organizational and community context: a framework for building evidence on dissemination and implementation in health services research. Adm Policy Ment Health. 2008;35(1-2):21-37.
19. Catho G, Centemero NS, Waldispühl Suter B, Vernaz N, Portela J, Da Silva S, et al. How to Develop and Implement a Computerized Decision Support System Integrated for Antimicrobial Stewardship? Experiences From Two Swiss Hospital Systems. Front Digit Health. 2020;2:583390.

20. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc. 2001;8(6):527-34.

21. Nicolini D, Powell J, Conville P, Martinez-Solano L. Managing knowledge in the healthcare sector. A review. International Journal of Management Reviews. 2008;10(3):245-63.

22. Paranjape K, Schinkel M, Hammer RD, Schouten B, Nannan Panday RS, Elbers PWG, et al. The Value of Artificial Intelligence in Laboratory Medicine. Am J Clin Pathol. 2021;155(6):823-31.

23. Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. Journal of Medical Artificial Intelligence. 2021;5.

24. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf. 2008;34(4):228-43.

25. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf. 2014;23(4):290-8.

# SUPPLEMENTARY APPENDIX

## Appendix 1

### *Topic list – implementation AI*

**Introduction**

*Do not record audio*

- Introduction of interviewers
- Aim of study
- Procedure interview

**Baseline**

*Do not record audio*

- Introduction of participant
- Sex
- Age
- Medical specialty
- Experience level

**TOPIC LIST**

*Start recording audio*

**Prior knowledge/opinion AI**

- Our definition of AI is: 'the use of complex algorithms and software to simulate human cognition.
- How would you describe your experience with AI?
- In what context was this
  - » Research
  - » Education
  - » Clinical practice
- What is your opinion on the use of AI in healthcare?
- What is the most important potential of AI in healthcare?
- What is the most important problem of AI in healthcare?

**Case vignette**

Start off with a short explanation of our prediction model, the context in which it can be used. Then explain that the case vignette will be used to gather the opinions about the use of our model, and that there are no right or wrong answers.

*Read case vignette to participant + hand the case vignette to participant on paper (see below)*

- What is your first impression of the case that was just drawn?
- Would you perform blood cultures for this patient?

    - » If yes: why?
    - » If no: when would you?

We developed a prediction model and implemented it in EPIC. It takes the values that are known for this patient, and predicts whether the blood culture will come back positive or not. For this patient the model predicted that the blood culture will come back negative, and you will be advised not to perform the blood culture.

- To what extent would/wouldn't you trust this prediction?
- Would you take this prediction into consideration regarding:
    - » Performing the blood culture
    - » Starting antibiotic treatment
- Would your way of thinking or acting be influenced if the prediction is presented differently? For example in percentages/risk groups/positive vs negative?
    - » What is needed to optimize this?



**Barriers**

The next questions are not specifically about the case, but about AI in healthcare in general.

- Which barriers do you foresee regarding working with an AI model in your clinical practice?
- What would be needed to remove these barriers?
- In case of multiple barriers: how do these barriers relate to each other/ prioritizing?
- What would be a worst case scenario when using an AI model in your clinical practice?

**Facilitators**

- Which factors would facilitate working with an AI model in your clinical practice?
- What would be needed to sustain these facilitating factors?
- In case of multiple facilitating factors: how do these barriers relate to each other/ prioritizing?
- What would be an ideal situation when using an AI model in your clinical practice?

**Implementation**

- What does the future with regards to AI in healthcare look like?
- Who could make use of AI models in clinical practice?
    - » Medical specialists only?
    - » Also other physicians; residents etc.?
    - » Also other healthcare providers; nurses, etc.?
    - » The patient?
- Who would be end responsible for the outcome when using AI in healthcare?
- What would be the most important factor for sustainable implementation of AI in healthcare?

**Closing question**

Is there anything that you would like to share, that has not been discussed yet?

**CASE VIGNETTE**

Age: 64

Sex: man

**Past medical history**

Type 2 Diabetes

Hypertension

COPD

**Medication use at home**

Metformin (3x 500mg)

Perindopril 1dd 8mg

Amlodipine 1dd 10mg

Spiriva 1dd

**History**

The patient experienced pain in the lower abdomen last night. During the night, the patient developed a fever of 39.5 degrees Celsius, and he collapsed when getting out of bed. The patient has no history of diarrhea or vomiting, has not experienced any chills, and he does not cough more than usual.

**Vital signs**

Heartrate: 106/min

Blood pressure: 115/76 mmHg

Respiratory rate: 21/min

Temperature: 38.9 degrees Celsius

Saturation: 93% on room air

EMV max

**Lab results**

CRP: 40

Leukocytes: 14

Sodium: 142

Potassium: 4.1

Creatinine: 98

**Microbiology**

No previous culture results are available.

Appendix 2

*AI in clinical practice*

# Implementing Artificial Intelligence in Clinical Practice



**Thank you for considering participating in this survey. It will take approximately 5 minutes to complete.**

## Participant information

**We kindly ask you to participate in a survey on the use of artificial intelligence (AI) in clinical practice. Researchers have developed many AI tools for healthcare, but few have made it into practice. We suspect the low adoption rates may result from the fact that the end-user (you) is rarely involved in these projects. In preceding interviews and focus group discussions, we identified several barriers and facilitators to implementing AI. With this survey, we aim to quantify further the importance of these barriers and facilitators on a larger scale.**

**1. Do you consent to participating in this survey?**

Yes

No

**2. Do you ever order or process blood cultures as part of your medical practice?**

Yes

No

**2. What's your age?**

Please select one category:

18-25
26-30
31-35
36-40
41-45
46-50
51-55
56-60
61-65
>65

Required Field

**3. What is your main specialty?**

Selecteer er één

Emergency Medicine

Intensive Care
Microbiology
Internal Mediine
Longgeneeskunde
Gastro-enterology
Reumatology
Urology
Geriatrics
Neurology
Surgery
Orthopaedics
Gynaecology
Other

Required Field

**4. How many years of experience do you have in that specialty (including residency)?**

Please type a number here:

Required Field

# Let's now talk about how your experience with Artificial Intelligence (AI)

## For our definition, we think that AI entails the use of complex algorithms and software to emulate human cognition.

**5. In what context have you ever come into contact with AI?**

Select all that apply (except for when the final answer is selected).

☐
I have used AI in clinical practice ☐

I have used AI in research ☐

I have developed an AI algorithm myself ☐

I am interested in AI and read about it ☐

I come into contact with AI in my daily life ☐

I have never come into contact with AI



**6. AI implementation in clinical practice could lead to cost effectiveness, improved workflows, and improved patient outcomes. Which of these benefits is most important to you?**

Rank the items in order of importance (1 = most important; 3 is least important)

1. Cost effectiveness; 2. Patient outcomes; 3. Workflow

1. Cost effectiveness; 2. Workflow; 3. Patient outcomes

1. Workflow; 2. Cost effectiveness; 3. Patient outcomes

1. Workflow; 2. Patient outcomes; 3. Cost effectiveness

1. Patient outcomes; 2. Cost effectiveness; 3. Workflow

1. Patient outcomes; 2. Workflow; 3. Cost effectiveness

**7. I would only trust an AI algorithm after I can first use it next to my own clinical judgement, to experience and judge the performance before fully adopting it.**

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

**8. I worry that AI algorithms will take over the enjoyable parts of my wok in the future.**

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

# The final questions will be specifically about our AI algorithm, which predicts the outcomes of blood cultures in the emergency department (ED)

**9. I feel the number of blood cultures we draw in the ED is:**

Very low

Low

Adequate

High

Very high

**10. What percentage of the blood cultures drawn in the ED in the Netherlands do you think will turn out to be positive (disregarding contamination)?**

Please provide a number between 0-100.

# Imagine you are consulting on a patient in the ED and decide to order a blood culture. Is that really needed for that specific patient?

## Our research group has developed an AI algorithm which predicts the outcomes of blood cultures drawn in the ED. The final questions are specifically about this scenario.

**11. I see the added value of using such an algorithm to better use blood culture testing in the ED.**

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

**12. I would like to learn about the algorithm and its use through:**

An optional explanation when using the algorithm in the electronic health records

Separate documents on the hospital's intranet

Presentations during handovers and educational meetings

A training period, in which the research team is present on the floor to answer questions

**13. I would like to see the algorithm's prediction for the individual patient as:**

Categories: low, intermediate, or high risk of a positive blood culture

Absolute risk: the percentage change of a positive blood culture

Binary outcome: the recommendation to do or withhold blood culture testing

## Appendix 3

### *Total output of ERIC strategies in Excel. ERIC, Expert Recommendations for Implementing Change*

| | ERIC Strategies | Cumulative Percent | Patient Outcomes & Quality | Relative Advantage | Acceptability | Trialability | Structural Characteristics | Tension for Change | Compatibility | Access to Knowledge & Information | Knowledge & Beliefs about the Intervention |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Identify and prepare champions | 283% | 44% | 42% | 22% | 12% | 27% | 43% | 21% | 24% | 40% |
| 3 | Conduct educational meetings | 258% | 47% | 21% | 12% | 8% | 5% | 17% | 10% | 79% | 36% |
| 4 | Promote adaptability | 237% | 5% | 24% | 91% | 22% | 21% | 17% | 47% | 7% | 10% |
| 5 | Access new funding | 234% | 15% | 21% | 11% | 35% | 9% | 5% | 9% | 7% | 36% |
| 6 | Conduct ongoing training | 224% | 41% | 24% | 11% | 0% | 14% | 4% | 41% | 10% | 12% |
| 7 | Capture and share local knowledge | 203% | 43% | 17% | 28% | 26% | 23% | 13% | 14% | 12% | 24% |
| 8 | Conduct local needs assessment | 201% | 3% | 24% | 36% | 10% | 18% | 43% | 21% | 3% | 24% |
| 9 | Inform local opinion leaders | 193% | 26% | 23% | 12% | 23% | 14% | 39% | 5% | 7% | 26% |
| 10 | Conduct cyclical small tests of change | 170% | 5% | 21% | 11% | 19% | 5% | 4% | 16% | 3% | 12% |
| 11 | Tailor strategies | 167% | 16% | 17% | 7% | 21% | 10% | 13% | 10% | 0% | 12% |
| 12 | Develop a learning collaborative | 154% | 15% | 7% | 20% | 17% | 10% | 4% | 14% | 14% | 14% |
| 13 | Facilitate relay of clinical data to providers | 153% | 22% | 21% | 21% | 12% | 23% | 13% | 19% | 12% | 20% |
| 14 | Develop educational materials | 152% | 28% | 14% | 12% | 0% | 0% | 0% | 5% | 53% | 36% |
| 15 | Distribute educational materials | 140% | 31% | 10% | 12% | 12% | 0% | 13% | 0% | 53% | 14% |
| 16 | Build a coalition | 127% | 0% | 14% | 12% | 15% | 27% | 9% | 21% | 3% | 10% |
| 17 | Facilitate | 124% | 0% | 10% | 27% | 21% | 9% | 0% | 21% | 15% | 20% |
| 18 | Conduct educational outreach visits | 124% | 41% | 10% | 10% | 4% | 5% | 0% | 5% | 24% | 10% |
| 19 | Visit other sites | 119% | 15% | 21% | 11% | 17% | 5% | 13% | 11% | 14% | 12% |
| 20 | Model and simulate change | 104% | 3% | 10% | 10% | 31% | 14% | 13% | 5% | 7% | 4% |
| 21 | Alter incentive/allowance structures | 97% | 3% | 28% | 0% | 0% | 18% | 22% | 10% | 0% | 14% |
| 22 | Stage implementation scale up | 92% | 0% | 10% | 0% | 27% | 14% | 4% | 10% | 7% | 20% |
| 23 | Develop a formal implementation blueprint | 89% | 0% | 7% | 8% | 19% | 18% | 13% | 5% | 14% | 4% |
| 24 | Organize clinician implementation team meetings | 86% | 5% | 14% | 10% | 4% | 14% | 4% | 14% | 14% | 4% |
| 25 | Use an implementation advisor | 86% | 8% | 7% | 8% | 24% | 5% | 13% | 18% | 17% | 6% |
| 26 | Provide local technical assistance | 78% | 5% | 0% | 4% | 13% | 18% | 0% | 14% | 24% | 6% |
| 27 | Audit and provide feedback | 78% | 18% | 10% | 4% | 13% | 5% | 17% | 7% | 3% | 4% |
| 28 | Provide ongoing consultation | 78% | 0% | 3% | 8% | 10% | 0% | 4% | 5% | 17% | 4% |
| 29 | Purposely reexamine the implementation | 70% | 0% | 7% | 12% | 13% | 0% | 0% | 26% | 0% | 4% |
| 30 | Facilitate relay of clinical data to providers | 70% | 0% | 10% | 4% | 4% | 0% | 22% | 5% | 10% | 12% |
| 31 | Involve patients/consumers and family members | 64% | 13% | 3% | 10% | 0% | 5% | 22% | 18% | 3% | 0% |
| 32 | Conduct ongoing training | 62% | 8% | 5% | 0% | 0% | 0% | 0% | 0% | 24% | 12% |
| 33 | Fund and contract for clinical innovation | 62% | 0% | 14% | 0% | 8% | 14% | 0% | 18% | 3% | 4% |
| 34 | Increase demand | 61% | 0% | 24% | 4% | 0% | 0% | 13% | 0% | 0% | 20% |
| 35 | Use advisory boards and workgroups | 60% | 8% | 10% | 4% | 8% | 5% | 13% | 5% | 0% | 8% |
| 36 | Develop academic partnerships | 60% | 22% | 0% | 0% | 0% | 5% | 0% | 0% | 10% | 12% |
| 37 | Shadow other experts | 59% | 5% | 3% | 12% | 4% | 5% | 0% | 5% | 21% | 4% |
| 38 | Use data experts | 57% | 8% | 10% | 8% | 17% | 0% | 17% | 0% | 0% | 4% |
| 39 | Provide financial incentives | 57% | 0% | 3% | 4% | 0% | 5% | 4% | 0% | 10% | 12% |
| 40 | Obtain and use patients/consumers and family feedback | 46% | 8% | 7% | 4% | 4% | 5% | 0% | 18% | 0% | 4% |
| 41 | Involve executive boards | 43% | 6% | 3% | 0% | 8% | 14% | 13% | 5% | 0% | 0% |
| 42 | Change physical structure and equipment | 42% | 0% | 3% | 0% | 0% | 32% | 0% | 7% | 0% | 0% |
| 43 | Develop and implement tools for quality monitoring | 41% | 0% | 7% | 0% | 17% | 5% | 9% | 5% | 0% | 0% |
| 44 | Work with educational institutions | 34% | 0% | 3% | 0% | 0% | 5% | 0% | 0% | 7% | 4% |
| 45 | Recruit, designate, and train leadership | 32% | 5% | 5% | 0% | 0% | 0% | 0% | 0% | 3% | 4% |

# PART III

## OPTIMIZING THE TREATMENT OF SEPSIS

# CHAPTER 7

## WHAT SEPSIS RESEARCHERS CAN LEARN FROM COVID-19

Michiel Schinkel, MD[1], Harjeet S. Virk, MD[2], Prabath W.B. Nanayakkara, MD, PhD, FRCP[1], Tom van der Poll, MD, PhD[2], W. Joost Wiersinga, MD, PhD, MBA[3]

[1]Section General and Acute Internal Medicine, Department of Internal Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC, location VU University Medical Center, Amsterdam, The Netherlands.
[2]Center for Experimental and Molecular Medicine (CEMM), Amsterdam UMC, location AMC, University of Amsterdam, Amsterdam, the Netherlands.
[3]Division of Infectious Diseases, Department of Medicine, Amsterdam UMC, location AMC, University of Amsterdam, Amsterdam, the Netherlands

*AJRCCM 2021*
*DOI: 10.1164/rccm.202010-4023LE on October 30, 2020*

Despite intensive research efforts, the search for new therapeutic options for sepsis has yielded no result [1]. However, the ongoing COVID-19 pandemic shows that effective therapeutic options for the distinct subgroup of viral sepsis due to SARS-CoV-2 infection, can be found within months [2]. What can sepsis researchers learn from the way COVID-19 is studied?

## HETEROGENEITY

In clinical practice, recognition of the wider sepsis syndrome can improve awareness and timely initiation of treatment. However, when looking for new therapeutic options in a research setting, this broad approach may be less desirable. One of the questionable tenets of sepsis research has been whether the host response in sepsis represents a "final common pathway" irrespective of the source of infection or causative pathogens [1]. This would justify looking at the broader sepsis population in research, with the added benefit of having larger study cohorts. However, most believe that the host response is just too complex and that a "final common pathway" may simply not exist [1]. The resultant heterogeneity within the sepsis population is therefore considered to be a major limiting factor in finding specific sepsis therapies [1, 3]. Extensive efforts have thus been made to reveal homogeneous sepsis subgroups [1, 3, 4].



Shared and distinct gene expression profiles are found when pulmonary and abdominal sepsis are compared [3]; suggesting that part of the heterogeneity in the sepsis population could be explained by the infection site or invading pathogen. Several other studies, that aim to find homogeneous sepsis subgroups through various methods, show different distributions of infectious aetiology across the newly formed subgroups, again implying that infecting organisms are associated with differences in the host response [3]. One study even states: "we examined only datasets of patients with bacterial sepsis at admission, because the clustering algorithms may otherwise have been overwhelmed by the differing host responses to different types of infections" [4].

In contrast to the many different causative microorganisms and arguably differing host responses in sepsis, early COVID-19 studies show comparable gene expression profiles in their populations, such as the upregulation of chemokines and neutrophils [2, 5]. This is possibly one of the key reasons why there have already been positive randomized trials with therapeutic options for COVID-19 [2]. Despite mixed results in sepsis trials, dexamethasone treatment resulted in lowering of 28-day mortality in COVID-19, particularly in patients who receive respiratory support [2]. Perhaps, focusing on a single site of infection or infective agent took away much of the heterogeneity. Researchers in the field of sepsis may learn from this and adapt current research paradigms and trial designs in such a way that stratification per infection type is possible and statistically meaningful.

## OUTCOME MEASURES

Outcome measures for sepsis clinical trials have been frequently discussed. Trials using novel therapeutic options have failed to demonstrate a benefit in general outcomes such as rates of ICU admission or mortality [1]. In 2005, the International Sepsis Forum (ISF) proposed that sepsis researchers should widen the breadth of outcome measures that are used in clinical trials [6]. Mortality is an attractive outcome measure, but other patient-centered benefits such as quality of

life and long-term morbidity should not be overlooked. The ISF colloquium provided additional, clinically relevant, possibilities to show benefits of a treatment [6]. Nevertheless, the literature on new therapeutics for sepsis continues to be dominated by the search for short-term mortality benefits.

For COVID-19, the World Health Organization (WHO) recognized that a core set of outcome measures was needed to investigate this new disease and compare outcomes globally. Experts who proposed the outcome measures for sepsis in 2005, also did so for COVID-19 in 2020 [7]. This time, a minimal common outcome measures set was used globally.

Another advantage of focussing on a more defined disease state, such as SARS-CoV-2 infection, in contrast to all-cause sepsis, is that site specific outcome measures can be used. For instance, the Murray score to assess lung injury [7], or diffusion capacity to assess pulmonary function [8] are valuable outcomes that could potentially be improved by certain treatments. Obviously, it does not make sense to assess pulmonary function as an outcome in all sepsis patients.



## GLOBAL COLLABORATION

Just weeks after the COVID-19 outbreak in Wuhan, China, the WHO coordinated a global research roadmap [9]. Experts from various fields agreed on key questions and strategies to accelerate research. The WHO launched a COVID-19 Data Platform to collect global data through a pre-defined case report form (CRF) [9]. When patient data was collected with this CRF anywhere around the world, the same variables were documented and the criteria for COVID-19 diagnosis (e.g. PCR or CT-scan) were available. The CRF was widely adopted and created a unique opportunity for global collaborative efforts, with minimal missing data or different inclusion criteria.

Furthermore, global genomic alliances are providing insights into how clinical and immunological manifestations of infection, and its natural variability, are governed by human genetics. In this case, global collaborations help find specific individuals prone or resistant to disease, who are especially interesting when trying to elucidate pathophysiological mechanisms.

Besides the use of a standardized data collection, COVID-19 research further profiled itself through the use of popular messaging platforms such as Slack [10]. In the United States, a group of researchers created a Slack forum to coordinate research projects across the country, providing yet another opportunity to have comparable study results.

## PITFALLS

The COVID-19 pandemic created much urgency with researchers worldwide. So far, we have outlined positive aspects of the COVID-19 research field that sepsis researchers can learn from (Table 1). Inevitably, this urgency also created pitfalls. The pressure to quickly perform and publish new studies led to acceptance of flexibility in protocols and trial design, shorter turnaround times for peer-review at medical journals and omission of extensive testing in preclinical animal models. Although these practices speed up the research process, one should be aware that they can also lower the standard of medical research, as is evident by the retraction of several papers in prominent medical journals over the past months.

**Table 1.** Key aspects of COVID-19 research that sepsis researchers can learn from.

| Aspect | Message |
|---|---|
| Heterogeneity | COVID-19 is more homogenous than sepsis, and that has probably been helpful with identifying effective treatments. Sepsis researchers should therefore consider smaller/more homogenous subgroups for study. |
| Outcome measures | Widespread use of core outcome sets facilitates comparison and pooling across studies. Examples of core outcomes [7]:<br>Organ dysfunction<br>Biochemical parameters<br>Radiological findings<br>Duration of intervention<br>Quality of life<br>Resource use<br>Examining homogenous subgroups facilitates additional outcome measures (e.g. severity of lung injury) that would not be relevant to an all-cause sepsis population |
| Global collaboration | Global data platforms with a standardized case report form can facilitate pooling of sepsis research<br>National or global coordination of large research projects can streamline sepsis research<br>Popular messaging platforms can be excellent tools to aid trial coordination |



## COMPETING INTERESTS

The authors declare that they have no competing interests related to this work.

## FUNDING

No funding was received for this work.

## CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## REFERENCES


1. Van Der Poll T, Van De Veerdonk FL, Scicluna BP, Netea MG (2017) The immunopathology of sepsis and potential therapeutic targets. Nat. Rev. Immunol. 17:407–420
2. Wiersinga WJ, Rhodes A, Cheng AC, et al (2020) Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA - J. Am. Med. Assoc. 324:782–793
3. Stanski NL, Wong HR (2020) Prognostic and predictive enrichment in sepsis. Nat. Rev. Nephrol. 16:20–31
4. Sweeney TE, Azad TD, Donato M, et al (2018) Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters HHS Public Access. Crit Care Med 46:915–925. https://doi.org/10.1097/CCM.0000000000003084
5. Gardinassi LG, Souza COS, Sales-Campos H, Fonseca SG (2020) Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. Front Immunol 11:1636. https://doi.org/10.3389/fimmu.2020.01636
6. Marshall JG, Vincent JL, Guyatt G, et al (2005) Outcome measures for clinical research in sepsis: A report of the 2nd Cambridge Colloquium of the International Sepsis Forum. Crit Care Med 33:1708–1716. https://doi.org/10.1097/01.CCM.0000174478.70338.03
7. Marshall JC, Murthy S, Diaz J, et al (2020) A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect. Dis. 20:e192–e197
8. Zhao Ymiao, Shang Ymin, Song Wbin, et al (2020) Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. EClinicalMedicine 0:100463. https://doi.org/10.1016/j.eclinm.2020.100463
9. World Health Organisation (2020) A Coordinated Global Research Roadmap: 2019 Novel Coronavirus
10. 'A completely new culture of doing research.' Coronavirus outbreak changes how scientists communicate | Science | AAAS. https://www-sciencemag-org.vu-nl.idm.oclc.org/news/2020/02/completely-new-culture-doing-research-coronavirus-outbreak-changes-how-scientists. Accessed 5 Aug 2020

# CHAPTER 8

## ASSOCIATION OF CLINICAL SUB-PHENOTYPES AND CLINICAL DETERIORATION IN COVID-19: FURTHER CLUSTER ANALYSES

Michiel Schinkel[1], Brent Appelman[1], Joe Butler[1], Alex Schuurman[1] and W. Joost Wiersinga[1,2]
on behalf of the COVID Predict Study Group

[1] Center of Experimental and Molecular Medicine (C.E.M.M.), Location Academic Medical Center, Amsterdam UMC, Amsterdam, The Netherlands.
[2] Division of Infectious Diseases, Department of Internal Medicine, Location AMC, Amsterdam UMC, Amsterdam, The Netherlands.

*Intensive Care Med (2021) 47:482–484*
*https://doi.org/10.1007/s00134-021-06363-9*

Dear Editor,

Coronavirus disease 2019 (COVID-19) presents in various ways [1]. Recently in the journal, Legrand and colleagues identified three distinct clinical sub-phenotypes of COVID-19, which may help recognize patients at high risk of deterioration [2]. Earlier work in sepsis has shown that clinical phenotypes may help understand the heterogeneity in disease presentation and inform trial design [3], [4]. The retrospective cohort study of Legrand et al. consisted of 893 patients of which 608 were used for cluster analysis, after excluding patients with missing data. Their thorough selection yielded 22 candidate variables for cluster analysis, including disease history, demographics, symptoms and concomitant medication.

We aimed to validate the findings by Legrand et al. in our Dutch CovidPredict cohort. This cohort consisted of COVID-19 positive patients admitted to ten teaching hospital across the Netherlands. COVID-19 was defined as a positive SARS-CoV-2 PCR or CORADS score of at least four [5]. Patients were included between 27 February and 4 December 2020. Approval was granted by the Institutional Review Board of the Amsterdam University Medical Centers (20.131).

We included 2019 patients and used similar candidate variables and number of clusters as Legrand et al [2] (see variables in the supplementary file). In total, 657 patients were treated in the intensive care unit (ICU) or died during the following 21 days of COVID-19. Three sub-phenotypes were identified, which are presented in Fig. 1 (see supplementary Table 1 and 2 for baseline characteristics, and Figure 1-3 for cluster characteristics).

**Sub-phenotype 1** (n=592) mainly included young (median age 63 [IQR = 53-74]) females (74.5%), characterized by a high prevalence of gastro-intestinal complaints (84.3%) and sputum production (63%). Comorbidities and medication usage were scarce. The composite outcome of ICU admittance/death rates was relatively low compared to the other groups (24.7%).

**Sub-phenotype 2** (n=876) included more males (80.4%) with a median age of 63 [IQR = 53-73.1] years, few comorbidities and the lowest medication usage of all three groups. Patients presented with less symptoms than those in sub-phenotype 1, but ICU admittance/death rates were higher (31.2%).

**Sub-phenotype 3** (n=551) mostly consisted of older (median age 76 [IQR = 69.1-81.1]) males (80.4%) with multiple comorbidities, mainly diabetes (62.4%), hypertension (87.7%) and other cardiovascular diseases (71.5%), and consequent medication usage. Patients reported less symptoms such as dyspnea (67%), headache (8.7%) and myalgia (11.6%). ICU admission and/or 21-day mortality occurred in 43.2% of patients.

In parallel with Legrand et al., sub-phenotype 1 was characterized by a large percentage of women and had the most favorable outcome. Sub-phenotype 3 differentiated itself by an older age together with a higher prevalence of comorbidities and a most unfavorable outcome.

The distributions of clinical characteristics were largely comparable to the original study across all sub-phenotypes. Notable differences with Legrand et al. were the relatively low age and percentages of women in sub-phenotype 2. We speculate that some female patients who were clustered as sub-phenotype 2 in the original study were clustered into sub-phenotype 1 in our study, perhaps due to slight differences in the prevalence of baseline characteristics in our more severely ill population. We believe the main value of these sub-phenotypes lies not with their

ability to discriminate between clinical outcomes, but in their potential to understand disease heterogeneity and find more homogeneous patient subgroups that may respond more similarly to certain treatments.

In conclusion, our large multicenter cohort of hospitalized COVID-19 patients showed largely similar distributions of the characteristics as Legrand et al. found, albeit in a more severely ill population. We validated the robustness of these three clinical phenotypes, which are strongly related to clinical outcomes.



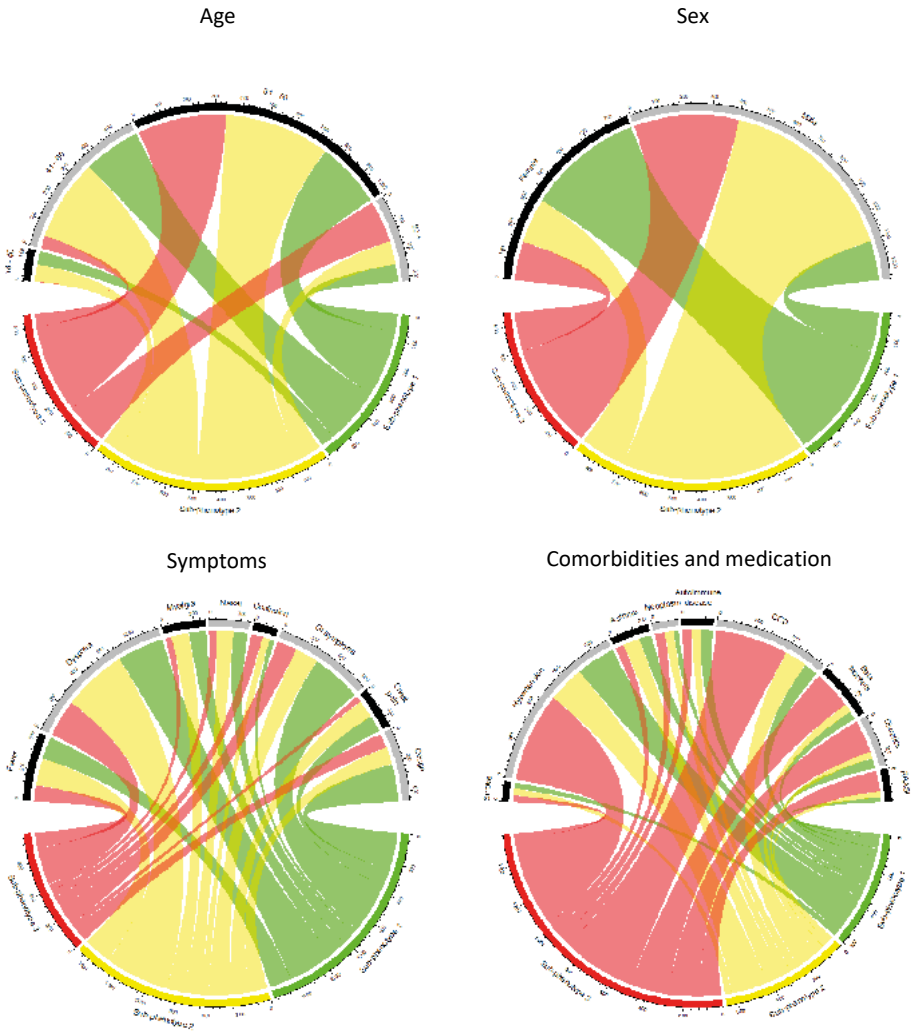



**Figure 1. Chord diagrams of the distributions of traits within the three sub-phenotypes in hospitalized patients with COVID-19.** In these chord diagrams the ribbons connect from the phenotype to the variables described. The proportion on the circle represents which group is more likely to have these characteristics traits (age, sex, symptoms, comorbidities and medication). Abbreviations used: GI = gastro-intestinal, CCD: Chronic Cardiovascular Disease, RAASi = Renin-Angiotensin-Aldosterone System Inhibitor.

## ACKNOWLEDGEMENTS

The members of the COVID Predict Study Group are: Renee A. Douma: Department of Internal Medicine, Flevoziekenhuis, Almere, The Netherlands. Joop P van den Berg: Department of Internal Medicine, VieCurie Medical Centre, Venlo, the Netherlands. Tom Dormans: Department of Intensive Care, Zuyderland Medical Center, Heerlen, The Netherlands. Auke C. Reidinga: Department of Intensive Care, Martiniziekenhuis, Groningen, The Netherlands. Niels C. Gritters van den Oever: Department of Intensive Care, Treant Zorggroep, Emmen, the Netherlands. Peter G. Noordzij: Department of Anesthesiology and Intensive Care, St Antonius Hospital, Nieuwegein, The Netherlands. Suat Simsek: Department of Internal Medicine, Northwest Clinical, Alkmaar, The Netherlands. Bart P. A. Spaetgens: Department of Internal Medicine, Maastricht University, Maastricht University Medical Center, Maastricht, The Netherlands. Paul W.G. Elbers: EDIC, Department of Intensive Care Medicine, Research VUmc Intensive Care (REVIVE), Amsterdam Medical Data science (AMDS), Amsterdam Cardiovascular Sciences (ACS), Amsterdam Infection and Immunity Institute (AI&II), Amsterdam UMC, location VUmc, Amsterdam, The Netherlands. Martijn Beudel: Department of Neurology, Amsterdam Neuroscience Institute, Amsterdam UMC, location AMC, Amsterdam, The Netherlands.

## REFERENCES


1. W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, and H. C. Prescott, "Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review," *JAMA - J. Am. Med. Assoc.*, vol. 324, no. 8, pp. 782–793, 2020.
2. Matthieu Legrand et al., "Differences in clinical deterioration among three sub-phenotypes of COVID-19 patients at the time of first positive test: results from a clustering analysis," *Intensive Care Med.*, pp. 15–17, 2020.
3. C. W. Seymour *et al.*, "Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis," *JAMA - J. Am. Med. Assoc.*, vol. 321, no. 20, pp. 2003–2017, 2019.
4. M. Schinkel, H. S. Virk, P. W. B. Nanayakkara, T. van der Poll, and W. J. Wiersinga, "What Sepsis Researchers Can Learn From COVID-19," *Am. J. Respir. Crit. Care Med.*, pp. 1–9, 2020.
5. M. Prokop *et al.*, "CO-RADS-A categorical CT assessment scheme for patients with suspected COVID-19: definition and evaluation Original research," *Radiology*, vol. Published, no. 1, 2020.

# SUPPLEMENTARY APPENDIX

## Methods

### *Descriptive statistics*

The study population was stratified based on cluster assignment and based on outcome (ICU and/or death). Group differences for all candidate variables were tested in univariate analyses. Differences in age, as a continuous variable, were tested with a Mann-Whitney U test or Kruskal Wallis test when appropriate, and presented with a median value and range. All other variables were categorical and differences were tested with a Pearson's Chi-square test. They are presented with counts and percentages.

### *Data collection*

The data was collected within the Covid Predict project, which is a multicenter cohort study in the Netherlands which aims to collect data on all admitted COVID-19 patients[1], [2]. All data is collected from the electronic health records.



### *Candidate variables*

The candidate variables for this cluster analysis were derived from the initial paper by Legrand et al. They selected 22 candidate variables, of which 20 were identically documented in our dataset. The final two variables in their analysis (heart disease and vascular disease) were documented with us in one single variable: cardiovascular disease. Therefore, we ended up using the following 21 variables in our clusters:

#### Demographics

- Sex, age and smoking status

#### Comorbidities

- Hypertension, diabetes, asthma, malignant neoplasms, autoimmune and immunodeficiency disorders and cardiovascular disease

#### Medications

- Beta-blockers, diuretics, RAAS inhibitors

#### Symptoms at admission

- Fever (>38.3 degrees Celsius), dyspnea, chest pain, nasal and/or throat symptoms, neurological symptoms, cough, myalgia, headache, gastro-intestinal symptoms (abdominal pain, diarrhea or vomiting)

### *Missingness*

Since only patients without missing values for the candidate variables could be included, only 2019 patient records out of the complete dataset of 2909 could be used for the cluster analyses.

## Cluster analysis

The cluster analyses were conducted using the R statistical software, version 3.6.1. In accordance with the Legrand paper[3], the distance between the samples was calculated using the Gower method, which accounts for the data type. A distance matrix was created using the daisy function from the cluster package[4]. This distance matrix was then used as input for the consensus cluster analysis with the ConsensusClusterPlus package[5]. Within the ConsensusClusterPlus function, we specified the use of the partitioning around medoids (PAM) clustering method, Pearson distance for the consensus matrix, resampling of both observations and features with 80% retention and 1,000 repetitions.

Since our main analysis focused on validating the clusters that were found by Legrand et al[3], we chose k=3 for the number of clusters to be presented. The consensus CDF plot and area under the CDF curve with increasing number of K suggest that this may also be the optimal solution in our dataset (see supplementary figure 1, 2 and 3).



### *Visualization*

The three clusters were visualized in a similar manner as in the Legrand paper. The Circlize package in R was used to create four chord diagrams for the age, sex, symptoms and concomitant medication/ comorbidity categories[6].

### *Limitations*

As the original study design was built around the missingness of variables in the Legrand paper there were various limitations. The most important limitation derives from the possible missed candidate variables. As our cohort consists out of admitted patients, missing data is scarce. Therefore our sub-phenotypes could have been represented by different candidate variables than in the original paper.

The second most important consideration comes from the fact that the Dutch healthcare system uses conservative do not resuscitation policies among the elderly with comorbidities. This limitation can affect the outcome of ICU admittance and/or death in specific subpopulations. This can be seen in sub-phenotype 3 where shared decision making lead to lower ICU admittance rates, due to their age andprevious medical conditions. This policy could have resultedin fewer ICU admittance in sub-phenotype 3.

Thirdly, a perfect replication of the Legrand paper was impossible, since our population only consisted of patients admitted to the hospital. This meant that the population was likely more severely ill. Although validation of the results in a different population is also a strength of this study, it may have led to some dissimilarities in the prevalence of the baseline characteristics. This can be seen, for example, in the higher median age across all sub-phenotypes. In the paper by Legrand and colleagues, there arguably is a slight, although not explicitly significant age difference between phenotype 1 and 2. The higher average age in our population shifts the complete age spectrum to the higher ages, which makes it so that little differences in the age trait are less visible in our population. An even more pronounced example may be the gender distributions in our sub-phenotypes. Specifically, percentage of women is even higher with regards to baseline in sub-

phenotype 1 and remarkably low in sub-phenotype 2 when compared with the original publication. We are not sure what the reason for this is. We speculate that some of the female patients that were clustered as sub-phenotype 2 in the Legrand study, have been clustered into sub-phenotype 1 in our study. This could potentially be the result of slight differences in the prevalence of symptoms and comorbidities across the board between the studies.

A final limitation of this analysis is that many policy changes have occurred in COVID-19 patient care over the past year. The paper by Legrand and colleagues only included patients seen between February 28$^{th}$ and March 26$^{th}$ of 2020, while our study included patients up until December 4$^{th}$. Arguably, patients seen in the second part of the year may have been somewhat different from those seen in March. However, since we show comparable results, we conclude that the impact of changes in carethroughout the year had little impact on this analysis.

### *Strengths*

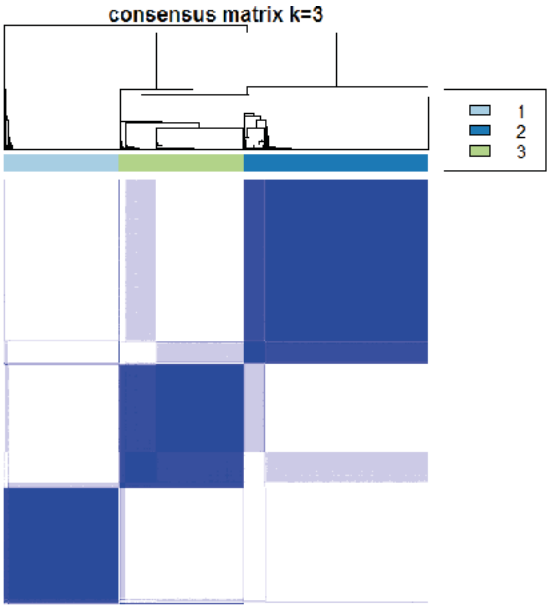
The main strength of this cluster analysis is that it showcases the usefulness and robustness of these types of analyses in medical practice. Although we performed the analysis in a more severely ill population in a different country with different policies, we still found similar patterns. Many studies have looked into clusters of patients in their own population, while few have set out to validate such findings in a different cohort. This study shows that cluster analyses can be robust and reproducible, which means that they can be used in a much wider setting.

**Supplementary table 1.** Baseline characteristics of the study population stratified per sub-phenotype.

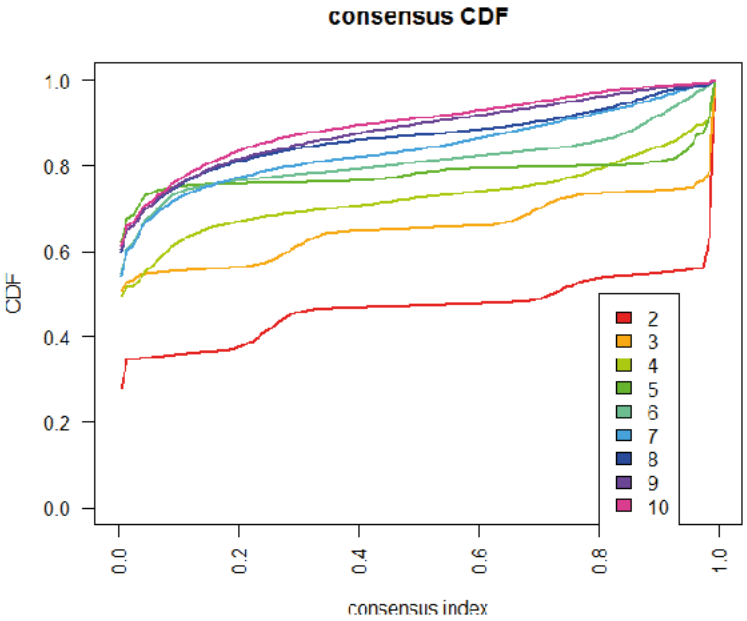
| | Sub-phenotype 1 | Sub-phenotype 2 | Sub-phenotype 3 | p-value |
|---|---|---|---|---|
| **n** | 592 | 876 | 551 | |
| **Sex = Female (%)** | 441 (74.5) | 172 (19.6) | 149 (27.0) | <0.001 |
| **Age (median [IQR])** | 63.00 [53.00, 74.00] | 63.00 [53.00, 73.10] | 76.00 [69.10, 81.10] | <0.001 |
| **Active smoker = Yes (%)** | 33 ( 5.6) | 38 ( 4.3) | 34 ( 6.2) | 0.281 |
| **Hypertension = Yes (%)** | 211 (35.6) | 198 (22.6) | 483 (87.7) | <0.001 |
| **Diabetes = Yes (%)** | 95 (16.0) | 72 ( 8.2) | 344 (62.4) | <0.001 |
| **Asthma = Yes (%)** | 65 (11.0) | 93 (10.6) | 47 ( 8.5) | 0.326 |
| **Malignant neoplasm = Yes (%)** | 29 ( 4.9) | 49 ( 5.6) | 54 ( 9.8) | 0.001 |
| **Auto-immune disorders = Yes (%)** | 65 (11.0) | 47 ( 5.4) | 54 ( 9.8) | <0.001 |
| **Chronic cardiovascular disease = Yes (%)** | 75 (12.7) | 122 (13.9) | 394 (71.5) | <0.001 |
| **Betablockers = Yes (%)** | 46 ( 7.8) | 46 ( 5.3) | 209 (37.9) | <0.001 |
| **Diuretics = Yes (%)** | 53 ( 9.0) | 37 ( 4.2) | 176 (31.9) | <0.001 |
| **RAAS inhibitors = Yes (%)** | 25 ( 4.2) | 32 ( 3.7) | 111 (20.1) | <0.001 |
| **Fever = Yes (%)** | 220 (37.2) | 273 (31.2) | 157 (28.5) | 0.005 |
| **Dyspnea = Yes (%)** | 461 (77.9) | 643 (73.4) | 369 (67.0) | <0.001 |
| **Chest pain = Yes (%)** | 166 (28.0) | 179 (20.4) | 56 (10.2) | <0.001 |
| **Nasal and throat symptoms = Yes (%)** | 145 (24.5) | 166 (18.9) | 72 (13.1) | <0.001 |
| **Neurological symptoms = Yes (%)** | 47 ( 7.9) | 77 ( 8.8) | 115 (20.9) | <0.001 |
| **Sputum = Yes (%)** | 373 (63.0) | 170 (19.4) | 145 (26.3) | <0.001 |
| **Myalgia = Yes (%)** | 169 (28.5) | 180 (20.5) | 64 (11.6) | <0.001 |
| **Headache = Yes (%)** | 194 (32.8) | 176 (20.1) | 48 ( 8.7) | <0.001 |
| **Gastro-intestinal symptoms = Yes (%)** | 499 (84.3) | 233 (26.6) | 200 (36.3) | <0.001 |
| **Do not resuscitate status = Yes (%)** | 144 (24.3) | 189 (21.6) | 288 (52.3) | <0.001 |
| **Do not intubate status = Yes (%)** | 116 (19.6) | 146 (16.7) | 264 (48.0) | <0.001 |
| **Death = Yes (%)** | 79 (13.3) | 95 (10.8) | 170 (30.9) | <0.001 |
| **ICU admission = Yes (%)** | 102 (17.2) | 213 (24.3) | 107 (19.4) | 0.003 |
| **ICU and/or Death (%)** | 146 (24.7) | 273 (31.2) | 238 (43.2) | <0.001 |

**Supplementary table 2.** Baseline characteristics of the study population stratified based on outcome.

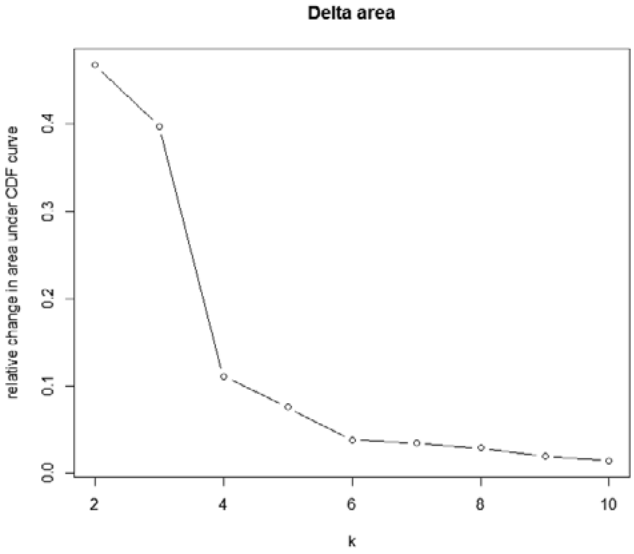
| | No ICU or Death | ICU and/or Death | p-value |
|---|---|---|---|
| **n** | 1362 | 657 | |
| **Sex = Female (%)** | 581 (42.7) | 181 (27.5) | <0.001 |
| **Age (median [IQR])** | 64.10 [54.00, 75.78] | 71.10 [62.00, 78.00] | <0.001 |
| **Active smoker = Yes (%)** | 70 ( 5.1) | 35 ( 5.3) | 0.943 |
| **Hypertension = Yes (%)** | 564 (41.4) | 328 (49.9) | <0.001 |
| **Diabetes = Yes (%)** | 307 (22.5) | 204 (31.1) | <0.001 |
| **Asthma = Yes (%)** | 142 (10.4) | 63 ( 9.6) | 0.614 |
| **Malignant neoplasm = Yes (%)** | 85 ( 6.2) | 47 ( 7.2) | 0.496 |
| **Auto-immune disorders = Yes (%)** | 115 ( 8.4) | 51 ( 7.8) | 0.663 |
| **Chronic cardiovascular disease = Yes (%)** | 358 (26.3) | 233 (35.5) | <0.001 |
| **Betablockers = Yes (%)** | 177 (13.0) | 124 (18.9) | 0.001 |
| **Diuretics = Yes (%)** | 164 (12.0) | 102 (15.5) | 0.036 |
| **RAAS inhibitors = Yes (%)** | 97 ( 7.1) | 71 (10.8) | 0.006 |
| **Fever = Yes (%)** | 406 (29.8) | 244 (37.1) | 0.001 |
| **Dyspnea = Yes (%)** | 962 (70.6) | 511 (77.8) | 0.001 |
| **Chest pain = Yes (%)** | 304 (22.3) | 97 (14.8) | <0.001 |
| **Nasal and throat symptoms = Yes (%)** | 280 (20.6) | 103 (15.7) | 0.010 |
| **Neurological symptoms = Yes (%)** | 135 ( 9.9) | 104 (15.8) | <0.001 |
| **Sputum = Yes (%)** | 460 (33.8) | 228 (34.7) | 0.717 |
| **Myalgia = Yes (%)** | 308 (22.6) | 105 (16.0) | 0.001 |
| **Headache = Yes (%)** | 323 (23.7) | 95 (14.5) | <0.001 |
| **Gastro-intestinal symptoms = Yes (%)** | 659 (48.4) | 273 (41.6) | 0.005 |




**Supplementary figure 1. Heatmap of the consensus matrix for K=3.** The consensus matrices are ordered by the consensus clustering which is depicted as a dendrogram atop the heatmap. The cluster memberships are marked by the colored rectangles between the dendrogram and heatmap.




**Supplementary figure 2. A plot of the Consensus Cumulative Distribution Function (CDF) specified by 100 bins for each k (clusters).** This figure shows for what number of k, the CDF reaches an approximate maximum and thereby consensus and cluster confidence is at a max for this k.




**Supplementary figure 3. A plot of the relative change in area under the Cumulative Distribution Function curve comparing K and K – 1.** This plot determines the relative increase in consensus and determine at whichk there is no appreciable increase. This figure presents values for K=2 through K=10.

## REFERENCES


1. Legrand. Differences in clinical deterioration among three sub-phenotypes of COVID-19 patients at the time of first positive test: results from a clustering analysis. *Intensive Care Med*. October 2020:1-3. doi:10.1007/s00134-020-06236-7
2. Maechler M, Struyf A, Hubert M, Hornik K, Studer M, Roudier P. Package 'cluster': Cluster Analysis Basics and Extensions. *R Top Doc*. 2015:79. https://cran.r-project.org/web/packages/cluster/cluster.pdf.
3. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn*. 2003;52(1-2):91-118. doi:10.1023/A:1023949509487
4. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811-2812. doi:10.1093/bioinformatics/btu393

# CHAPTER 9

## TIMELINESS OF ANTIBIOTICS FOR PATIENTS WITH SEPSIS AND SEPTIC SHOCK

Michiel Schinkel, MD[1,2], Rishi S. Nannan Panday, MD, PhD[1],
W. Joost Wiersinga, MD, PhD, MBA[2], Prabath W.B. Nanayakkara, MD, PhD, FRCP[1]

[1]Section Acute Medicine, Department of Internal Medicine, Amsterdam UMC, location VU University Medical Center. De Boelelaan 1118, 1081HZ, Amsterdam, The Netherlands.
[2]Section Infectious Diseases, Department of Internal Medicine, Amsterdam UMC, location Academic Medical Center. Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands.

*Journal of Thoracic Diseases 2020*
*doi: 10.21037/jtd.2019.10.35*

## INTRODUCTION

Sepsis is one of the major health problems of the 21st century. It is currently defined as a dysregulated host response to an infection, which causes life-threatening organ dysfunction and a mortality risk of about 10% (1). The risk of mortality increases to over 40% for patients with septic shock (1). Although there still is no specific treatment for sepsis, general treatment with fluids and antibiotics has been the gold standard for many years. A frequently cited paper by Kumar et al. in 2006 showed that every hour of delay in the administration of antibiotics decreased the chances of survival by 7.6% (2). Although the study by Kumar and colleagues was based on retrospective data and only counted the delay in antibiotics from the onset of persistent hypotension, the term 'golden hour of sepsis' was introduced, suggesting that there is only a small window of opportunity to optimize the treatment strategy for these patients. Since then, most treatment protocols for sepsis have focused on administering antibiotics as soon as possible. While this practice may benefit some patients, for others it might have detrimental consequences.

## SURVIVING SEPSIS CAMPAIGN GUIDELINES

Currently, the Surviving Sepsis Campaign (SSC) guidelines are widely used to guide treatment for patients with sepsis (3). The main focus of these guidelines is early identification of sepsis, treatment with broad spectrum antibiotics and administration of intravenous fluids when needed. Since the initiation of the SSC in 2002, the guidelines have proposed several bundles that included elements of treatment which have to be started within a specific time period. With newer iterations of the guidelines, the timeframe in which antibiotic treatment had to be initiated was shortened, without a high level of evidence for these updated recommendations (4–6). Following the 3-hour and 6-hour timeframes of the previous bundles, the latest update of the SSC guidelines proposed an "Hour-1" bundle to initiate treatment as early as possible for all patients suspected of having sepsis (7). This bundle was immediately challenged by many physicians. After extensive debates (8–10), the Society of Critical Care Medicine (SCCM) and the American College of Emergency Physicians (ACEP) finally issued a statement recommending against the use of the SSC 1-hour bundle, leaving many physicians and hospitals in doubt about which guidelines to use for patients with suspected sepsis in the emergency care setting.

## OVERUSE OF ANTIBIOTICS

The increasingly shortened timeframes in which guidelines recommend administration of antibiotics for sepsis have forced emergency care personnel to sacrifice diagnostic accuracy for speed (8). Limiting the time to perform a proper diagnostic work up has inevitably encouraged overuse of antibiotics (6). A study in the Netherlands showed that up to 43% of patients admitted to the intensive care unit because of sepsis were unlikely to even have an infection (11). Another study showed that 29% of patients who were diagnosed with sepsis and received antibiotics in the emergency department were unlikely to have an underlying bacterial infection (12). This unnecessary use of antibiotics can have many negative effects such as an increased rate of *Clostridium difficile* infections, organ injury and a disruption of the gut microbiome (5,13). On

a population level, overuse of antibiotics can increase antibiotic resistance, leading to a further acceleration of this global crisis (14). On the other hand, it is questionable whether this practice actually benefits all patients with sepsis.

## EVIDENCE FOR EARLY ADMINISTRATION OF ANTIBIOTICS

Following the paper by Kumar and colleagues (2), numerous studies on the effects of early administration of antibiotics for patients with sepsis have been conducted. A systematic review and meta-analysis by Sterling et al. in 2015 included 11 retrospective observational studies on this subject (15). Although there was significant heterogeneity between the included studies, the authors concluded that there was no significant increase in risk of mortality for each hour of delay in treatment, when looking at the pooled effect of these studies. Another two key retrospective studies have been published since. Both Seymour et al. and Liu et al. found significant increases in mortality for each hour of delay in antibiotics administration (16,17). This was most prominent for patients with septic shock. However, it should be acknowledged that multiple studies that found significant and often linear effects on mortality, favoring early administration of antibiotics, have limitations associated with their study design. Firstly, all these studies have been conducted retrospectively on databases that were not created for this purpose (5). Then, the outcomes have been adjusted for many variables, raising the risk of overadjustment (18), while often neglecting factors such as concomitant treatments, appropriateness of antibiotic therapy or confounding by indication (5,6). Lastly, the premise of a linear increase in mortality when antibiotic treatment is delayed is questionable (5). Time zero, or the time when the infection or organ dysfunction started, is hard to define. This could have been hours to even days before the presentation in the emergency department. It thus seems highly unlikely that the first few hours in the emergency department will see such an increase in mortality (5).



Besides retrospective analyses, there have also been some prospective studies on this subject. In 2012, Hranjec et al. evaluated the effects of conservative initiation of antimicrobial treatment, rather than aggressive and early administration of antibiotics for critically ill surgical intensive care unit patients with suspected infection (19). The authors concluded that an aggressive approach significantly increased the risk of mortality when compared with a conservative approach. Also, the conservative approach led to more appropriate antimicrobial therapy and a shorter treatment period. De Groot and colleagues published another study that prospectively evaluated early administration of antibiotics, which did not show any benefits of this practice (20). Finally, in 2018, the first and thus far only randomized trial on the subject of early antibiotics for sepsis was conducted by our group: the prehospital antibiotics against sepsis (PHANTASi) trial (21). This large trial evaluated the effects of administration of antibiotics to patients with sepsis in the ambulance, rather than in the ED. Emergency medical personnel was trained to recognize patients with sepsis. Afterwards patients were randomized to receive either usual supportive care or a dose of 2000 mg ceftriaxone in addition to the supportive care in the ambulance. The usual care group received their first dose of antibiotics in the ED. The early intervention resulted in a difference in time to antibiotics of 96 minutes between the intervention and usual care group. However, the 28- and 90-day mortality

rates did not differ between the groups. The only difference that was found between these groups, was the 28-day readmission rate, which was significantly higher in the control group (7% vs 10%). The population of patients with septic shock was just 3% of the complete study population, which made it hard to detect potential effects of early antibiotics on mortality in this subgroup.

The PHANTASi trial provided a couple of interesting findings. The design of the trial gave it the unique opportunity to randomize between early and late antibiotic treatment, which would otherwise have been unethical given the standard practice at that point in time. Some important limitations of this study should be addressed. Firstly, this was a select population of patients that had a suspected infection and a minimum of two of a selection of three of the systemic inflammatory response syndrome (SIRS) criteria (temperature > 38°C or < 36°C, heart rate > 90 beats per minute, respiratory rate > 20 per minute) which was the gold standard for diagnosing sepsis at the time the study was conducted. As just 3% of the study population had septic shock, we cannot compare these results to other studies given that these included only critically ill sepsis patients. However, the patient-mix in the PHANTASi trial was probably very similar to the general emergency department population of sepsis patients (22). Secondly, the reduction in time to antibiotics was just 96 minutes. Over 40% of patients in the usual care group received antibiotics within one hour of presentation to the ED and about 85% of patients within 3 hours (21). Even in the retrospective studies that report significant increases in mortality when antibiotics are not administered early, the risk of mortality does not increase immensely in these first hours. It is thus questionable whether we can expect any significant differences in this short time frame. Lastly, we have to consider the fact that, although there was no difference in mortality rates, there was a difference in readmission rates. It has been proposed that the early administration of antibiotics may have inhibited the development of organ dysfunction in some patients (23). There may thus well be a beneficial effect of early administration of antibiotics in selected groups (23).



Summarizing the evidence on the early administration of antibiotics for patients with sepsis, we can conclude that evidence for supporting this practice mainly comes from retrospective observational studies, with all the limitations attached. One prospective study even found favorable effects from a conservative approach regarding initiation of antimicrobial therapy (19). Furthermore, when significant effects favoring early administration of antibiotics are found, this is usually in the most critically ill patients with septic shock. The most compelling evidence, from the only randomized trial (the PHANTASi trial) on this subject (21), does not show a mortality benefit from early administration of antibiotics in a population as often seen in the emergency department.

## IDENTIFYING PATIENTS WITH SEPSIS

When considering the appropriateness of existing sepsis protocols which focus on early administration of antibiotics, we also have to examine the specific groups of patients who are labelled as having sepsis. Currently, according to the Sepsis-3 guidelines, sepsis should be suspected in patients who have a positive quick Sequential Organ Failure Assessment (qSOFA) score and have an increase in the Sequential Organ Failure Assessment (SOFA) score of 2 or more points, due to suspected infection (1). To break it down, the definition consists of two components: organ dysfunction

quantified by the qSOFA and SOFA score and suspicion of infection. Both these components cause problems when used to select patients to treat with antibiotics in an early stage. Firstly, some parts of the SOFA score are based on the results of laboratory test, which are not immediately available in every setting. The SOFA score is therefore rarely used outside the intensive care unit (ICU) and the use of SOFA in conjunction with qSOFA to define sepsis is thus rather confusing and impractical. In clinical practice, scores such as the qSOFA, National Early Warning Score (NEWS) or SIRS are often used as independent tools to detect patients with a high risk of mortality due to suspected infection (24). They are easy to use, but far from accurate (24). The qSOFA is not sensitive enough to be used as a screening tool (25,26), while the SIRS criteria lack specificity and cause many false positive results (24). With current protocols, physicians could be forced to either underdiagnose a substantial amount of patients with sepsis, or treat a significant proportion of these patients with antibiotics, while many may not need them.

The other part of the definition of sepsis states that the organ dysfunction has to be caused by a suspected infection. This is purely based on clinical judgement, as there are no objective criteria for this component of the definition. More experienced clinicians will likely be more accurate when suspecting an infection. Increasing the accuracy with which physicians can assess the likelihood of an infection will increase the validity of the sepsis criteria and may also improve the accuracy of scores like qSOFA and SIRS. Assessment of patients with sepsis by a senior attending would greatly help in this regard. Furthermore, it is of importance that only patients who are suspected of having a bacterial infection, and not viral infection, are treated with antibiotics. The study by Minderhoud et al. showed that out of a total of 78 patients (29%) who received antibiotics without evidence of a bacterial infection, 21 patients (8%) actually suffered from proven or suspected viral infections (12).



## CONSIDERATIONS

We have discussed the intricacies of the assessment and treatment of patients suspected of having sepsis in a non-ICU setting. It remains challenging to accurately suspect infection and identify patients with sepsis, especially in the elderly with atypical presentations (27). Even more difficult perhaps is the distinction between bacterial and non-bacterial disease, for which there are no reliable diagnostic tests yet. Treating a general group of patients who are suspected of having sepsis, causes many patients to be treated with antibiotics unnecessarily. Protocols that have challenged physicians to sacrifice diagnostic accuracy in order to initiate treatment within a certain timeframe, have only amplified this effect. Furthermore, these protocols could also be misused as a performance measurement for hospitals, with unwanted consequences. As there is little evidence to support the early administration of antibiotics, especially for the general emergency department population of patients with sepsis, an updated international guideline is needed. A striking fact about the current situation is that we have had the same problem with the management of community-acquired pneumonia and do not seem to have learned from that experience. A quality measure was instituted in 2002 in the United States, forcing physicians to treat patients with suspected pneumonia with antibiotics within four hours (28). This practice, not based on high quality evidence, led to the same problem of overdiagnosis and unnecessary use of antibiotics. Eventually, the negative effects were recognized and the quality metric was removed.

Considering all the available evidence on this subject, it seems reasonable to suggest that rapid administration of empiric antibiotics will benefit critically ill sepsis patients with signs of shock benefit (16,17) and that there is certainly no margin for error in this group (29). However, for patients who are suspected of having a systemic infection, but who are not in shock, physicians could take additional time to gather information to further confirm the diagnosis of sepsis and the suspicion of a bacterial cause. This is even more relevant given the technological advances regarding molecular diagnostic tests such as polymerase chain reaction (PRC) to rapidly detect causative agents with high sensitivity (30). Instead of being challenged to treat patients within a set period of time, physicians should be challenged to identify the patients with suspected sepsis who will not be hurt by taking time to gather additional patient data and only administer antibiotics when it really could benefit the patient. New guidelines for the treatment of patients with sepsis should thus not only stipulate goals separately for patients with sepsis and patients with septic shock, they should also avoid pursuing specific time periods in which treatment should be initiated for the general sepsis population. However, physicians should be encouraged to perform an adequate work-up as soon as possible.

The clinical dilemma between early administration of antibiotics according to the guidelines and an approach more similar to what we have just described was presented in the New England Journal of Medicine by Mi and colleagues (31). In this case vignette of two patients with suspected infection, arguments were made for both immediate administration of antibiotics and a more careful approach where additional information could be gathered before deciding to administer antibiotics (31). Interestingly, a poll at the end of the article showed that the total of 3118 responders were split fifty-fifty between these two options. Readers of this article seem to value the existing literature differently. Another possibility would be that many readers chose their answer based on the existing guidelines, not having had the time to evaluate the literature themselves. Consensus about the evidence and updated international protocols are much needed, to make sure that sepsis care is based on the best available evidence and is comparable between different hospitals.



## CONCLUSION

Studies regarding the use of early antibiotics for patients with sepsis are often limited by problems inherent to this heterogeneous and enigmatic syndrome. With the existing guidelines, physicians are challenged to treat patients suspected of having sepsis within a very short period of time, while the real challenge should be to identify patients who would not be harmed by withholding treatment with antibiotics until the diagnosis of infection with a bacterial origin is confirmed and the appropriateness of a course of antibiotics can be evaluated more adequately. Therefore, in the general population of patients with sepsis, taking the time to gather additional data to confirm the diagnosis should be encouraged without a specific timeframe, although physicians should be encouraged to perform an adequate work-up as soon as possible. Patients with suspected sepsis and signs of shock should immediately be treated with antibiotics, as there is no margin for error.

- In the general population of patients with sepsis, taking the time to gather additional data to confirm the diagnosis should be encouraged without a specific timeframe, although physicians should be encouraged to perform an adequate work-up as soon as possible.
- Critically ill patients with suspected sepsis and signs of shock should be treated with antibiotics as soon as possible, as there is no margin for error with these patients.

**Box 1. key recommendations**

## FINANCIAL SUPPORT

No financial support was received for this work.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES


1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA [Internet]. 2016 Feb 23 [cited 2019 Mar 4];315(8):801. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.0287
2. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Crit Care Med [Internet]. 2006 Jun [cited 2019 Apr 15];34(6):1589–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16625125
3. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign. Crit Care Med [Internet]. 2017 Mar [cited 2019 Apr 18];45(3):486–552. Available from: http://insights.ovid.com/crossref?an=00003246-201703000-00015
4. Kalantari A, Rezaie SR. Challenging the one-hour sepsis bundle. Vol. 20, Western Journal of Emergency Medicine. eScholarship; 2019. p. 185–90.
5. Singer M. Antibiotics for sepsis: Does each hour really count, or is it incestuous amplification? Vol. 196, American Journal of Respiratory and Critical Care Medicine. American Thoracic Society; 2017. p. 800–2.
6. Klompas M, Calandra T, Singer M. Antibiotics for Sepsis - Finding the Equilibrium. Vol. 320, JAMA - Journal of the American Medical Association. American Medical Association; 2018. p. 1433–4.
7. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. Intensive Care Med [Internet]. 2018 [cited 2019 Jul 19];44:925–8. Available from: https://doi.org/10.1007/s00134-018-5085-0
8. Marik PE, Farkas JD, Spiegel R, Weingart S, Aberegg S, Beck-Esmay J, et al. POINT: Should the Surviving Sepsis Campaign Guidelines Be Retired? Yes. Vol. 155, Chest. Elsevier Inc; 2019. p. 12–4.
9. Marik PE, Farkas JD, Spiegel R, Weingart S, Aberegg S, Beck-Esmay J, et al. Rebuttal From Drs Marik, Farkas, Spiegel et al. Vol. 155, Chest. Elsevier Inc; 2019. p. 17–8.
10. Levy MM, Rhodes A, Evans LE, Antonelli M, Bailey H, Kesecioglu J, et al. Rebuttal From Drs Levy, Rhodes, and Evans. Vol. 155, Chest. Elsevier Inc; 2019. p. 19–20.
11. Klein Klouwenberg PMC, Cremer OL, van Vught LA, Ong DSY, Frencken JF, Schultz MJ, et al. Likelihood of infection in patients with presumed sepsis at the time of intensive care unit admission: A cohort study. Crit Care. 2015 Sep 7;19(1).
12. Minderhoud TC, Spruyt C, Huisman S, Oskam E, Schuit SCE, Levin MD. Microbiological outcomes and antibiotic overuse in Emergency Department patients with suspected sepsis. Neth J Med [Internet]. 2017 Jun [cited 2019 Sep 12];75(5):196–203. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28653945
13. Hiensch R, Poeran J, Saunders-Hao P, Adams V, Powell CA, Glasser A, et al. Impact of an electronic sepsis initiative on antibiotic use and health care facility–onset Clostridium difficile infection rates. Am J Infect Control. 2017 Oct 1;45(10):1091–100.
14. WHO | Antimicrobial resistance. WHO. 2019;
15. Sterling SA, Miller R, Pryor J, Puskarich MA, Jones AE. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis HHS Public Access. Crit Care Med [Internet]. 2015 [cited 2019 Jul 15];43(9):1907–15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4597314/pdf/nihms688736.pdf
16. Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med [Internet]. 2017 [cited 2019 Sep 12];196(7):856–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28345952
17. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med [Internet]. 2017 [cited 2019 Sep 12];376(23):2235–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28528569

18. Jager KJ, Zoccali C, MacLeod A, Dekker FW. Confounding: What it is and how to deal with it. Kidney Int. 2008;73(3):256–60.

19. Hranjec T, Rosenberger LH, Swenson B, Metzger R, Flohr TR, Politano AD, et al. Aggressive versus conservative initiation of antimicrobial treatment in critically ill surgical patients with suspected intensive-care-unit-acquired infection: A quasi-experimental, before and after observational cohort study. Lancet Infect Dis. 2012 Oct;12(10):774–80.

20. De Groot B, Ansems A, Gerling DH, Rijpsma D, Van Amstel P, Linzel D, et al. The association between time to antibiotics and relevant clinical outcomes in emergency department patients with various stages of sepsis: a prospective multi-center study. 2015;

21. Alam N, Oskam E, Stassen PM, Exter P van, van de Ven PM, Haak HR, et al. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. Lancet Respir Med [Internet]. 2018 Jan [cited 2019 Apr 18];6(1):40–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29196046

22. Groenewoudt M, Roest AA, Leijten FMM, Stassen PM. Septic patients arriving with emergency medical services: a seriously ill population. Eur J Emerg Med. 2014 Oct 1;21(5):330–5.

23. Quinten VM, van Meurs M, Ligtenberg JJ, Ter Maaten JC. Prehospital antibiotics for sepsis: beyond mortality? Lancet Respir Med [Internet]. 2018 Mar 1 [cited 2019 Apr 23];6(3):e8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29508708

24. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the Emergency Department. American Journal of Emergency Medicine. 2018;

25. Fernando SM, Tran A, Taljaard M, Cheng W, Rochwerg B, Seely AJE, et al. Prognostic Accuracy of the Quick Sequential Organ Failure Assessment for Mortality in Patients With Suspected Infection. Ann Intern Med [Internet]. 2018 Feb 20 [cited 2019 Oct 14];168(4):266. Available from: http://annals.org/article.aspx?doi=10.7326/M17-2820

26. Van Der Woude SW, Van Doormaal FF, Hutten BA, Nellen FJ, Holleman F. Classifying sepsis patients in the emergency department using SIRS, qSOFA or MEWS.

27. Caterino JM, Kline DM, Leininger R, Southerland LT, Carpenter CR, Baugh CW, et al. Nonspecific Symptoms Lack Diagnostic Accuracy for Infection in Older Patients in the Emergency Department. J Am Geriatr Soc [Internet]. 2019 Mar [cited 2019 Sep 23];67(3):484–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30467825

28. Spiegel R, Farkas JD, Rola P, Kenny J-E, Olusanya S, Marik PE, et al. The 2018 Surviving Sepsis Campaign's Treatment Bundle: When Guidelines Outpace the Evidence Supporting Their Use. Ann Emerg Med [Internet]. 2019 Apr [cited 2019 Aug 21];73(4):356–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30193754

29. Gilbert DN, Kalil AC, Klompas M, Masur H, Winslow DL. IDSA POSITION STATEMENT: Why IDSA Did Not Endorse the Surviving Sepsis Campaign Guidelines. Clin Infect Dis [Internet]. 2017 [cited 2019 Sep 12];45:486. Available from: http://academic.oup.com/cid/advance-article/doi/10.1093/cid/cix997/4645414

30. Nannan Panday RS, Wang S, Van De Ven PM, Hekker TAM, Alam N, Nanayakkara PWB. Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. PLoS One. 2019 Mar 1;14(3).

31. Mi MY, Klompas M, Evans L. Early administration of antibiotics for suspected sepsis. N Engl J Med. 2019 Feb 7;380(6):593–6.

# CHAPTER 10

## TOWARDS UNDERSTANDING THE EFFECTIVE USE OF ANTIBIOTICS FOR SEPSIS

M. Schinkel, MD[1, 2*], K. Paranjape, Msc[1, 3*], J. Kundert, Msc[3],
R.S. Nannan Panday, PhD[1, 4], N. Alam, PhD[1], P.W.B. Nanayakkara, FRCP[1]

[1] Section General and Acute Internal Medicine, Department of Internal Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC, location VU University Medical Center. De Boelelaan 1118, 1081HZ, Amsterdam, The Netherlands.
[2] Center of Experimental and Molecular Medicine (C.E.M.M.), Amsterdam UMC, location Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.
[3] Roche Diagnostics Corporation, 9115 Hague Road, Indianapolis, IN 46256, United States.
[4] Center for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
* Both authors contributed equally to the manuscript.

*Chest 2021*
*https://doi.org/10.1016/j.chest.2021.04.038*

# ABSTRACT

## Background

The benefits of early antibiotics for sepsis have recently been questioned. Evidence for this mainly comes from observational studies. The only randomized trial on this subject, the PHANTASi trial, did not find significant mortality benefits from early antibiotics. It is still plausible that subgroups of patients benefit from this practice, given the heterogeneous nature of sepsis.

## Research Questions

Do subgroups of sepsis patients experience 28-day mortality benefits from early administration of antibiotics in a prehospital setting? And what key traits drive these benefits?

## Study Design and Methods

We used machine learning to conduct exploratory partitioning cluster analysis to identify possible subgroups of sepsis patients who may benefit from early antibiotics. We further tested the influence of several traits within these subgroups using a logistic regression model.

## Results

We found a significant interaction between age and benefits of early antibiotics (p=0.03). When we adjusted for this interaction and several other confounders, there was a significant benefit of early antibiotic treatment (OR = 0.07; 95%-CI = 0.01-0.79; p = 0.03).

## Interpretation

An interaction between age and benefits of early antibiotics for sepsis has not been reported before. When validated, it can have major implications for clinical practice. This new insight into benefits of early antibiotic treatment for younger sepsis patients may enable more effective care.



## Keywords

Antibiotics; Sepsis; Age; Machine Learning; PHANTASi trial; Mortality; Prehospital

**Study question**

Are there specific subgroups of sepsis patients who are likely to benefit from early antibiotic treatment?

**Results**

We found a significant interaction between age and benefits of early antibiotics, associating early treatment with a significant decrease in 28-day mortality among younger sepsis patients.

**Interpretation**

Our results suggest that we should immediately consider antibiotic treatment in younger patients, while early treatment does not seem to have much beneficial effects in older sepsis patients.

**Take-home Points**

## INTRODUCTION

Sepsis is a major health problem worldwide. A recent study estimated the global incidence of sepsis to be nearly 50 million cases annually with 11 million sepsis-related deaths[1]. Dysregulation of the host response to infections can cause organ dysfunction and subsequently leads to these high mortality rates[2]. Sepsis is a truly heterogeneous syndrome[3,4], caused by different pathogens at various sites (e.g. respiratory tract, urinary tract, or abdominal), which makes it difficult to develop general guidelines that will benefit all sepsis patients.

Researchers have aimed to identify specific subgroups of sepsis patients in order to tailor the treatment. Seymour and colleagues, for example, categorized four clinical sepsis phenotypes with similar traits, that may also respond similarly to certain treatments[5]. Current sepsis treatment mainly includes administration of antibiotics and intravenous fluids. The subcategorization of sepsis patients could help use these options more effectively when given to the right patient at the right time.

Most patients suspected of having systemic infections rapidly receive antibiotic treatment in the emergency department (ED). There is a long-standing belief that every hour of delay in administration of antibiotics leads to an increased risk of mortality, as suggested by Kumar et al. in 2006[6]. Many treatment protocols for sepsis have been guided by this belief, ultimately resulting in an international effort called the Surviving Sepsis Campaign (SSC) guideline 1-hour bundle[7].

Recently the benefits of early antibiotic treatment in all patients with suspected sepsis have been questioned[8–11]. Physicians are forced to sacrifice diagnostic accuracy, in order to treat these patients early, which contributes to overuse of antibiotics[8,12,13]. A Dutch study reported that 29% of suspected sepsis patients in the ED were unlikely to even have an infection[12]. In a recent review, we evaluated the literature on the benefits of early antibiotics for sepsis and concluded that the evidence for this is mainly derived from observational studies[8]. The only randomized controlled trial on this subject, called the Prehospital Antibiotics Against Sepsis (PHANTASi) trial, conducted by our research group, did not show significant benefits of early antibiotic treatment in a pre-hospital setting[14].

Although there is no conclusive evidence supporting the early use of antibiotics in all patients with suspected sepsis, it is plausible that subgroups of patients may benefit from early antibiotic treatment. In this study, we aim to identify subgroups of patients in the PHANTASi trial cohort who are likely to benefit from early antibiotic treatment and study their key traits using machine learning[15].

## STUDY DESIGN AND METHODS

### Database

The PHANTASi trial database was used for this study[14]. The PHANTASi trial randomized 2672 patients with suspected sepsis to either receive antibiotic treatment in the ambulance (intervention) or antibiotic treatment once the patient had arrived in the ED (control). This resulted in a median difference in time to antibiotics of 96 minutes (IQR: 36-128) between the groups. The study ran between June 2014 and June 2016. Patients were included when they were at least 18 years of age, were suspected of having an infection, and had at least two Systemic Inflammatory Response Syndrome (SIRS) criteria, with a mandatory temperature ≥38°C or ≤36°C. The original trial was registered at ClinicalTrials.gov, number NCT01988428. More details on this study can be found here[14,16].

Vital parameters and laboratory results were recorded in the ambulance and in the ED. Any treatments, including an early dose of antibiotics in the ambulance in the intervention group, were recorded. Diagnoses were confirmed by an expert panel and sepsis severity was categorized according to the 2001 international sepsis criteria[17], which were the gold standard at the time. The study was powered to detect differences in the primary outcome, which was 28-day mortality[14].



### Statistical Analysis

Statistical analyses were performed in R 3.5[18], and in R modules within the Alteryx software (Alteryx Inc, Irvine CA, USA)[19], which is an extraction transformation and loading application. Differences between non-normally distributed and continuous variables were assessed with a Mann-Whitney U test[20]. Differences between categorical variables were tested with a chi-square test. Normality of the data was assessed with histograms and Q-Q plots. A two-tailed p-value of $<0.05$ was considered to be statistically significant.

Machine learning algorithms were used to conduct exploratory partitioning cluster analysis to identify possible factors impacting the benefits of early antibiotic treatment. This clustering approach involved three broad phases: exploratory data analysis, preliminary cluster diagnostics, and then focused cluster partitioning based on key traits.

During the exploratory data analysis, unsupervised machine learning techniques (K-means, K-medians, and Neural Gas clustering) were performed in order to identify any relevant cluster patterns exhibited by combinations of traits with either known or suspected associations with 28-day mortality. Twenty-two exploratory analyses were performed involving various traits (outlined in e-Table 1: Exploratory K-Centroids Diagnostic Data Mining Trials). These clusters assessed various clinical factors obtained in the ambulance, ED, as well as deterioration between ambulance and ED

(delta in particular traits such as heart rate, respiratory rate, etc.). We visually assessed each cluster pattern outcome to gain general insight and help shape the direction of subsequent, more focused, clustering techniques.

We identified three specific focused clustering combinations, outlined in Table 1, for further evaluation and subsequent cluster diagnostics, based specifically on clinical factors obtained in the ambulance. A thorough pre-assessment K-Centroid diagnostic analysis was performed for these specific combinations of key traits. This involved identifying possible traits that could have a strong cluster relationship, and then algorithmically evaluating the mathematically ideal number of clusters (k) for each combination. Cluster diagnostic results, including supporting Adjusted Rand (ARI) and Calinski-Harabasz (CH) indices for each selected k-value, are represented in Table 1. The ARI was used to help provide a measure of agreement, or similarity, between partitions; the CH provided a measure for separation and inter-cluster density. The assessment process evaluated the suitable number of clusters (k) by maximizing ARI and CH, when compared to k alternatives, in order to increase cluster performance and quality. Once the number of clusters was determined for each possible trait combination, the clustering assignment was attempted and associated to each patient record. We used K-Means clustering for each grouping and no additional unit standardization was applied to input fields. See Table 1 for further details. These cluster analyses focused primarily on better understanding previously unknown relationships within the data, as well as to help focus the direction of subsequent, more traditional, multivariable logistic regression statistical analysis.



To further test associations between 28-day mortality and various traits, a multivariable logistic regression model was used. The raw model was adjusted for confounders using the 10% change-in-estimate criterion, as is one of the accepted methods of confounder identification[21,22]. Also, full models with all a priori identified theoretical confounders are presented[23].

In some cases, age was not used as a continuous variable, but as a dichotomous variable, The categories were created by splitting the dataset in the 50% youngest and 50% oldest patients, in order to obtain equally large numbers of patients in both groups[22]. The age ranges in these groups were 18 - 75 and 76 - 100 years respectively.

## RESULTS

### Exploratory partitioning cluster analysis

Clusters of similar patients were created based on various patient characteristics and with the use of various unsupervised machine Learning techniques. Based on the most favorable Rand index values, a K-means cluster algorithm based on age, heart rate in the ambulance, and temperature in the ambulance was selected to generate two clusters (mean ARI: 0.93; mean CH: 4485.1). The patterns produced using this model consistently resulted in strong ties associated with the age trait, seen in figure 1, with partitioning occurring around the age of 70. Figure 1 illustrates three different two-dimensional representations of the same clusters, generated based on age, heart rate, and temperature. Though these are simplified representations of the three-dimensional clusters, they clearly show that the age trait is the most important driver of the clusters.

**Table 1.** K-Centroids Cluster Diagnostics

| K-Centroids Method | Min/Max Cluster Parameter | Number of Traits Evaluated | Traits Assessed | Number of Clusters (k) for Partitioning | Diagnostic Results | | Cluster Results | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Adjusted Rand (Mean) | Calinski-Harabasz (Mean) | Cluster | Size | Average Distance | Max Distance | Separation |
| K-means | 2/8 | 6 | Heart Rate (Ambulance); Systolic BP (Ambulance); Diastolic BP (Ambulance); Respiratory Rate (Ambulance); Temperature (Ambulance); Blood Oxygen Saturation (Ambulance) | 3 | 0.61 | 342.11 | 1 | 1290 | 99.28 | 2691.1 | 34.4 |
| | | | | | | | 2 | 54 | 135.46 | 2694.8 | 982.3 |
| | | | | | | | 3 | 1175 | 53.94 | 1016 | 33.3 |
| K-means | 2/8 | 2 | Heart Rate (Ambulance); Temperature (Ambulance) | 5 | 0.80 | 5266.8 | 1 | 734 | 5.92 | 13.5 | 8.87 |
| | | | | | | | 2 | 865 | 3.34 | 8.02 | 7.66 |
| | | | | | | | 3 | 182 | 6.88 | 31.27 | 11.33 |
| | | | | | | | 4 | 130 | 10.31 | 58.13 | 14.49 |
| | | | | | | | 5 | 608 | 4.2 | 10.84 | 8.12 |
| K-means | 2/8 | 3 | Age; Heart Rate (Ambulance); Temperature (Ambulance) | 2 | 0.93 | 4485.1 | 1 | 1671 | 5.29 | 19.34 | 12.43 |
| | | | | | | | 2 | 848 | 8.59 | 39.58 | 11.66 |

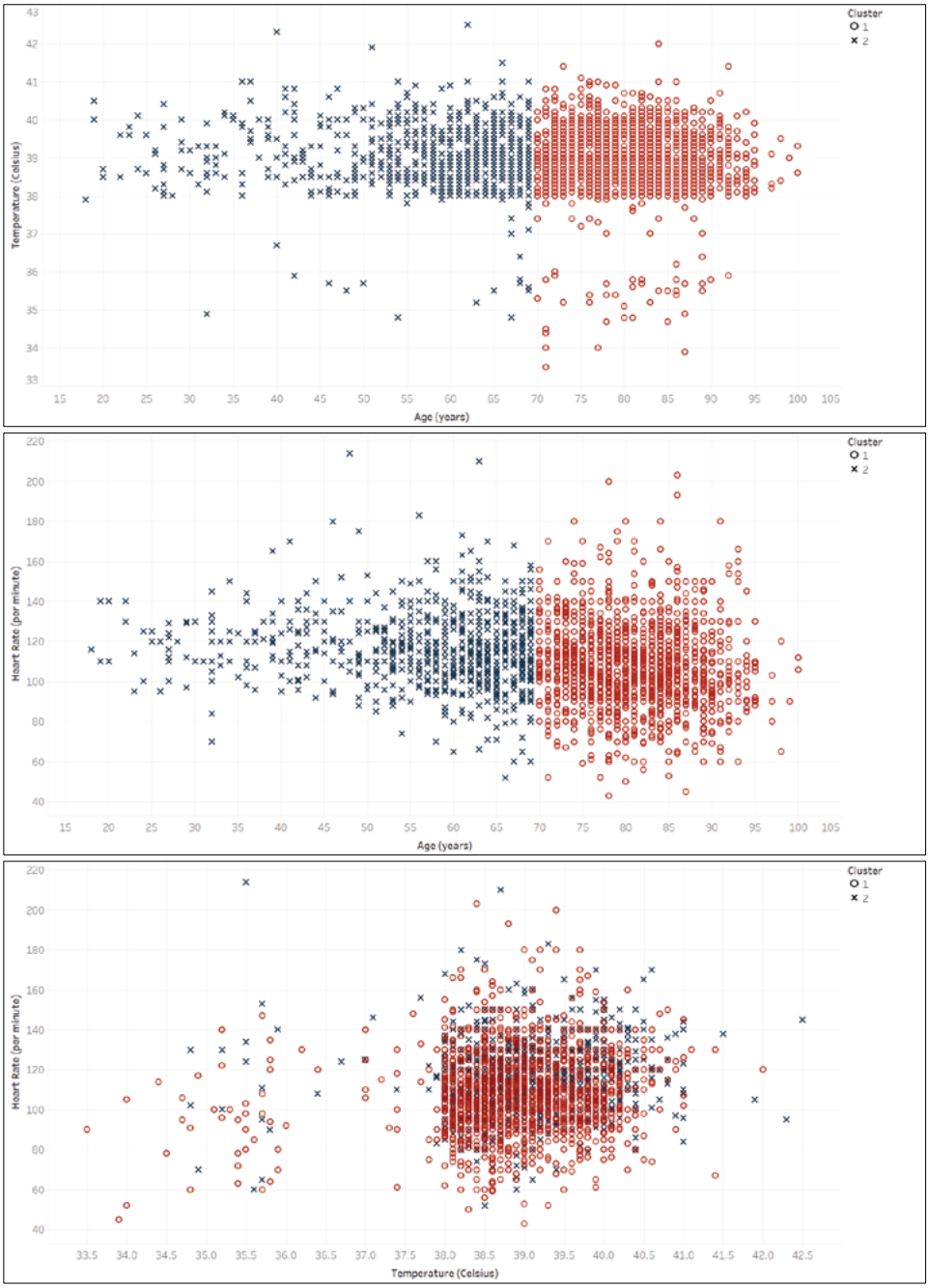


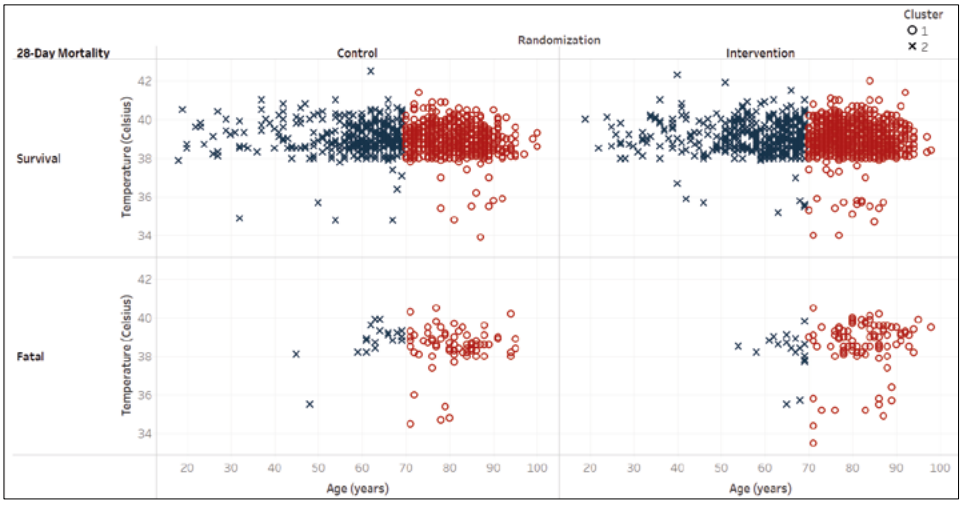

**Figure 1. Three two-dimensional visualizations of the same clusters with k-means clustering based on age, heart rate and temperature.**

In figure 2a, patients were categorized based on designated cluster and separated by randomization group and 28-day mortality outcome. For simplicity, we opted to only present a two-dimensional representation in this figure, since further insights are mostly derived from the age axis. The figure identifies the control group (antibiotics administered in the ED) from the intervention group (antibiotics in the ambulance), and separates patients who survived after 28 days from those deceased. Cluster 1 (denoted: O) resulted in 1671 patients with a mean age of 80.6. Cluster 2 (denoted: X) produced 848 patients with a mean age of 57.5. There were also 153 patients categorized as outliers based on inconclusive clinical factors and were not assigned a cluster. Additional analysis yields that younger patients seen in cluster 2 may exhibit a slight lowering of the overall 28-day mortality rate in the intervention group (4.0%) when compared to younger patients in the control group (5.0%), while this is less pronounced in cluster 1 with older patients. Mortality rate percentages associated with each cluster are further outlined in figure 2b.

## Logistic regression modelling

We created an association model to quantify the initial finding of a possible interaction between age and the effect of early antibiotic treatment. We used a logistic regression model to explain






Randomization / 28-Day Mortality

| Cluster | Control | | Intervention | |
|---|---|---|---|---|
| | Survival | Fatal | Survival | Fatal |
| 1 | 89.61% | 10.39% | 90.58% | 9.42% |
| | 621 | 72 | 885 | 92 |
| 2 | 94.97% | 5.03% | 95.96% | 4.04% |
| | 359 | 19 | 451 | 19 |

**Figure 2a. Visualization of clusters with k-means clustering based on age and heart rate (with temperature as the third clustering variable) segmented by intervention status and mortality outcome.** 2b –. Mortality rate summary percentages with k-means clustering based on age, heart rate and temperature segmented by intervention status.

28-day mortality in all patients who were categorized as having sepsis (n=2617). This number differs from the complete population (n=2672), because some patients had diagnoses other than sepsis in retrospect. Baseline characteristics of the included patients are presented in Table 2.

We used 28-day mortality as dependent variable and intervention with early antibiotics (yes/no) as the main independent variable in our model. We also added the interaction between intervention and age (as a continuous variable) in the raw model, since this was the effect modifier we aimed

**Table 2.** Baseline characteristics of the complete sepsis population.

| | Control (N=1113) | Intervention (N=1504) | Total (N=2617) | p value |
|---|---|---|---|---|
| **Age, years** | | | | 0.509 |
| Median (IQR) | 75.0 (65.0, 83.0) | 76.0 (66.0, 83.0) | 76.0 (65.0, 83.0) | |
| **Sex** | | | | 0.763 |
| Male | 638 (57%) | 871 (58%) | 1509 (58%) | |
| Female | 475 (43%) | 633 (42%) | 1108 (42%) | |
| **Youngest or oldest half of the patients** | | | | 0.536 |
| Under 76 years | 559 (50%) | 737 (49%) | 1296 (50%) | |
| 76 years or above | 554 (50%) | 767 (51%) | 1321 (50%) | |
| **Sepsis severity** | | | | 0.341 |
| Non-severe Sepsis | 424 (38%) | 576 (38%) | 1000 (38%) | |
| Severe Sepsis | 653 (59%) | 863 (57%) | 1516 (58%) | |
| Septic shock | 36 (3%) | 65 (4%) | 101 (4%) | |
| **Charslon Comorbidity Index** | | | | 0.988 |
| Median (IQR) | 1.0 (1.0, 3.0) | 1.0 (0.0, 3.0) | 1.0 (1.0, 3.0) | |
| **Do not resuscitate order** | | | | 0.307 |
| No | 666 (61%) | 862 (59%) | 1528 (60%) | |
| Yes | 425 (39%) | 598 (41%) | 1023 (40%) | |
| **quick Sequential Organ Failure Assessment Score (qSOFA)** | | | | 0.003 |
| 2 or more | 176 (17%) | 310 (22%) | 486 (20%) | |
| Smaller than 2 | 855 (83%) | 1109 (78%) | 1964 (80%) | |
| **Use of immunosuppressiva medication** | | | | 0.799 |
| No | 960 (86%) | 1292 (86%) | 2252 (86%) | |
| Yes | 153 (14%) | 212 (14%) | 365 (14%) | |
| **Patient already on oral antibiotics before randomisation** | | | | 0.241 |
| No | 864 (79%) | 1189 (81%) | 2053 (80%) | |
| Yes | 224 (21%) | 274 (19%) | 498 (20%) | |
| **Pathogen resistant to ceftriaxone** | | | | 0.015 |
| Sensitive | 1106 (100%) | 1483 (99%) | 2589 (100%) | |
| Resistant | 0 (0%) | 8 (1%) | 8 (0%) | |
| **Blood culture results from ambulance/emergency department** | | | | < 0.001 |
| Negative | 829 (75%) | 1239 (83%) | 2068 (80%) | |
| Positive | 277 (25%) | 252 (17%) | 529 (20%) | |
| **28-day mortality** | | | | 0.753 |
| Survived | 1021 (92%) | 1386 (92%) | 2407 (92%) | |
| Died | 91 (8%) | 118 (8%) | 209 (8%) | |

to study. In the raw model, the effect of the intervention on 28-day mortality (OR = 0.13; 95%-CI = 0.02-1.10; p = 0.061) as well as the interaction term between age and the benefit of the intervention (OR = 1.03; 95%-CI = 1.00-1.05; p = 0.066) did not meet traditional measures of clinical significance. We then adjusted the model for a priori selected potential confounders, based on the 10% change-in-estimate criterion. This resulted in an adjustment based on qSOFA score and Charlson comorbidity index, after which other variables did not meaningfully change this adjusted model. The adjusted model showed a significant benefit of the intervention on 28-day mortality (OR = 0.07; 95%-CI = 0.01-0.79; p = 0.03) as well as a significant interaction term between age and the benefit of the intervention (OR = 1.03; 95%-CI = 1.00-1.06; p = 0.03). Additionally, we created a full model based on all a priori selected potential confounders, irrespective of their influence in this dataset. This approach has been proposed in the literature and provided similar results as the adjusted model, as can be seen in Table 3, which also shows the full list of variables that we had selected as possible confounders.

## Age as a categorical value

In the initial model, we used age as a continuous variable. Since we cannot be sure that the beneficial effects of early antibiotics decrease linearly with increasing age, we also created a model based on age groups. The age groups were created by a split based on the median age. This resulted in a cut off at the age of 76. The raw model, with age as dichotomous variable, did not show significant benefits of the intervention (OR = 0.68; 95%-CI = 0.02-1.10; p = 0.126), or interaction term between age and the benefit of the intervention (OR = 1.65; 95%-CI = 0.90-3.05; p = 0.110). We then adjusted the model for the same variables as the adjusted model in the previous analysis, and noticed that differences in the benefits of early antibiotics (OR = 0.63 95%-CI = 0.36-1.06; p = 0.082), just as



**Table 3.** Associations of various traits with 28-day mortality through logistic regression modelling

| | Age continuous | | | | | |
|---|---|---|---|---|---|---|
| | Raw model | | Adjusted model | | Full model | |
| Characteristics | OR (95% CI) | P | OR (95% CI) | P | OR (95% CI) | P |
| **Intervention (Y)** | 0.13 (0.02-1.10) | 0.061 | 0.07 (0.01-0.79) | 0.031 | 0.07 (0.01-0.80) | 0.031 |
| **Age** | 1.03 (1.01-1.05) | 0.001 | 1.03 (1.01-1.05) | 0.008 | 1.00 (0.99-1.03) | 0.583 |
| **Age * intervention** | 1.03 (1.00-1.05) | 0.066 | 1.03 (1.00-1.06) | 0.033 | 1.03 (1.00-1.07) | 0.030 |
| **Sex (F)** | | | | | 0.91 (0.66-1.24) | 0.543 |
| **Charlson comorbidity index (per point increase)** | | | 1.17 (1.09-1.25) | 0.001 | 1.12 (1.04-1.20) | 0.002 |
| **qSOFA (lower than 2)** | | | 0.46 (0.33-0.63) | 0.001 | 0.56 (0.40-0.78) | <0.001 |
| **Do not resuscitate order (Y)** | | | | | 3.75 (2.58-5.55) | <0.001 |
| **Antibiotics prior to hospital visit (Y)** | | | | | 1.34 (0.93-1.91) | 0.111 |
| **Immunosuppressive comedication (Y)** | | | | | 1.48 (1.00-2.16) | 0.046 |
| **Positive blood culture (Y)** | | | | | 1.37 (0.95-1.96) | 0.088 |
| **Ceftriaxone resistant pathogen (Y)** | | | | | 2.83 (0.38-14.00) | 0.235 |

the interaction term between age and the benefit of the intervention (OR = 1.89; 95%-CI = 0.99-3.63; p = 0.055) did not meet traditional measures of clinical significance. The full model, adjusted a priori with identified possible confounders, showed a similar benefit of early antibiotics as with age as a continuous variable (OR = 0.59; 95%-CI = 0.34-1.05; p = 0.063) and the interaction term between age and the benefit of the intervention also presented similar results (OR = 2.17; 95%-CI = 1.11-4.30; p = 0.025). See Table 3 for further details.

### Different cut-off values for age groups

In the analysis which used age as a dichotomous variable, we chose to split the groups based on the median age. Supplementary Table 2 presents results for other cut-off values. Many cut-off values between 75 and 83 years of age showed significant results.

## DISCUSSION

We re-evaluated the PHANTASi trial cohort to identify subgroups of patients who may benefit from early antibiotic treatment and the traits driving these subgroups. We found a significant interaction between age and intervention with early antibiotics, associating early antibiotic treatment with a significant decrease in 28-day mortality among younger patients. We showed that there is a significant interaction between age and the effect of early antibiotic treatment on mortality (p=0.04). When we adjusted for this interaction, along with other potential confounders, there was a significant association between intervention with early antibiotics and 28-day mortality (OR = 0.07; 95%-CI = 0.007-0.75; p = 0.03).



| Age dichotomous | | | | | |
|---|---|---|---|---|---|
| Raw | | Adjusted model | | Full Model | |
| OR (95% CI) | P | OR (95% CI) | P | OR (95% CI) | P |
| 0.68 (0.02-1.10) | 0.126 | 0.63 (0.36-1.06) | 0.082 | 0.59 (0.34-1.03) | 0.063 |
| 1.77 (1.14-2.77) | 0.012 | 1.60 (1.00-2.59) | 0.053 | 0.90 (0.54-1.51) | 0.679 |
| 1.65 (0.90-3.05) | 0.110 | 1.89 (0.99-3.63) | 0.055 | 2.17 (1.11-4.30) | 0.025 |
| | | | | 0.92 (0.67-1.26) | 0.613 |
| | | 1.18 (1.10-1.26) | <0.001 | 1.12 (1.04-1.20) | 0.003 |
| | | 0.45 (0.33-0.62) | <0.001 | 0.55 (0.39-0.77) | <0.001 |
| | | | | 4.17 (2.88-6.14) | <0.001 |
| | | | | 1.32 (0.91-1.88) | 0.132 |
| | | | | 1.46 (0.98-2.13) | 0.056 |
| | | | | 1.38 (0.95-1.97) | 0.084 |
| | | | | 2.55 (0.33-13.35) | 0.230 |

## In context

The three largest observational studies which evaluate the effect of time of antibiotic administration on mortality, have not assessed the interaction between the age of the patients and the benefits of early antibiotic treatment[24–26]. Over the past year, our research group has received several inquiries about the non-significant, but notably low relative risk of mortality in the younger patients in the original PHANTASi trial, which spiked our interest in finding subgroups of patients who may have benefitted from early antibiotics. We opted to start this study by performing exploratory partitioning cluster analysis, rather than focusing specifically on age, since this allowed us to provide a broader view of potential patient factors that could be associated with benefits of early antibiotics treatment. However, we soon found that age seemed to be the most important driver of clusters and that we needed to focus on this trait.

## Residual confounding

We tested the robustness of our results by using age as a continuous as well as a dichotomous variable, as well as using empirical and theoretical criteria to select the confounders we adjusted for. We thereby hoped to have limited residual confounding which is inherent to secondary analyses. Since this study is based on secondary analyses, p-values are difficult to interpret. The original study was not designed to detect this interaction, which makes it hard to find statistically significant results. We therefore focused on evaluating whether our findings remained similar when we examined different subgroups or adjusted the model for different potential confounders, while still providing p-values and confidence intervals for clarity.



We showed that the interaction between age and the intervention with early antibiotics was independent of the cut-off value we used for the age groups. In supplementary Table 2, we report p-values for the interaction between age and intervention for cut-off levels between the age of 70 and 85, which are significant at multiple thresholds. The absence of significant results at the lower and higher ends of that range is likely a reflection of the low numbers of patients and events in one of the two groups in those situations. This can also explain why the relative risk in the original publication of the PHANTASi trial did not reach statistical significance. The cut-off in the original publication was 65, which is a commonly accepted cut-off to define younger and older patients, but created a younger group (n=600) that was considerably smaller than the elderly group (n=2017).

## Clinical value

The interaction between age and benefits of early antibiotic treatment, which is associated with significant improvements in 28-day mortality in younger sepsis patients, can be clinically relevant. Knowing in which subcategory of patients benefits of early antibiotic treatment can be expected, will enable effective and optimized care.

Our results suggest that we should immediately consider antibiotic treatment in younger patients, while early treatment does not seem to have much beneficial effects in older sepsis patients. We do not propose a specific age cut-off for the benefits of early antibiotics, but we do believe that additional time to do a proper work-up may be taken with elderly sepsis patients, to confirm the diagnosis before initiating antibiotic treatment. This is especially helpful since diagnosing

sepsis in the elderly is often more challenging due to non-specific presentations[27]. Recent research indicates that early administration of antibiotics is associated with higher mortality when given to patients with greater diagnostic uncertainty[28]. Arguably, the diagnostic uncertainty may be higher in elderly patients, given the non-specific presentations. This provides an additional argument for withholding antibiotic treatment until the diagnosis is clearer.

We should note that our study only included patients with symptoms of sepsis. It may well be that early administration of antibiotics for elderly sepsis patients in practice is even less desirable, since this practice may even harm the patients with less specific presentations. Furthermore, there was only a small decrease in time to antibiotics (96 minutes) by intervening with antibiotics in the ambulance in this trial. In many settings, administration of antibiotics in the ambulance will result in larger decreases in time to antibiotics, which is possibly associated with an even stronger mortality benefit.

## Strengths

We examined an interaction which to our knowledge has never been reported before. The interaction between age and benefits of early antibiotic treatment may explain part of the variance in benefits of early antibiotic treatment which is observed throughout the literature on this subject[3,29]. Furthermore, we used data from the single randomized trial on this subject, which lowers the chance of residual. Lastly, we could evaluate the effect of potential confounders such as antibiotic sensitivities, while most studies on this subject lack this important data to evaluate adequacy of antibiotic treatments[30].



## Limitations

We recognize the limitations of performing secondary analyses. Subgroup effects can be misleading and can be explained by chance[31]. To minimize the risk that we found these results by chance, we performed several different analyses to see whether our results were robust. A second limitation is that we were not able to validate our findings in a similar cohort, since the PHANTASi trial was the only randomized trial on this subject and was conducted in a very specific setting. Validation of our findings in existing large observational cohorts could provide additional strength to our findings. However, such cohorts carry high risk of residual confounding and will not be able to undeniably validate or disprove our findings. A definite answer to whether young patients benefit from early antibiotics can only be given by another randomized study such as the PHANTASi trial.

## Interpretation

In conclusion, we have re-examined the effects of early antibiotic treatment for sepsis, finding a significant interaction between age and mortality benefits of this practice. Young sepsis patients seem to experience a significant mortality benefit from early antibiotic treatment in the ambulance, which reduces as age increases. This interaction has not been reported before. Validation studies in other cohorts are needed to confirm our findings, which could lead to a shift in the way we think about the pathophysiology of sepsis and the most optimal treatment strategies.

## ACKNOWLEDGMENTS

MS, KP, RSNP, NA and PWBN conceived the study. NA and RSNP were responsible for the database and PWBN was responsible for study supervision. MS, KP, JK and PWBN analyzed and interpreted that data. MS, KP, JK, RSNP and PWBN drafted the manuscript. All authors read, revised, and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## FINANCIAL SUPPORT

No financial support was received for this work.

## REFERENCES


1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet*. 2020;395(10219):200-211. doi:10.1016/S0140-6736(19)32989-7
2. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801. doi:10.1001/jama.2016.0287
3. de Grooth HJ, Postema J, Loer SA, Parienti JJ, Oudemans-van Straaten HM, Girbes AR. Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. *Intensive Care Med*. 2018;44(3):311-322. doi:10.1007/s00134-018-5134-8
4. Schinkel M, Virk HS, Nanayakkara PWB, van der Poll T, Wiersinga WJ. What Sepsis Researchers Can Learn from COVID-19. *Am J Respir Crit Care Med*. 2021;203(1):125-127. doi:10.1164/rccm.202010-4023LE
5. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019;321(20):2003. doi:10.1001/jama.2019.5791
6. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med*. 2006;34(6):1589-1596. doi:10.1097/01.CCM.0000217961.75225.E9
7. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med*. 2017;43:304-377. doi:10.1007/s00134-017-4683-6
8. Schinkel M, Nannan Panday RS, Wiersinga WJ, Nanayakkara PWB. Timeliness of antibiotics for patients with sepsis and septic shock. *J Thorac Dis Publ Ahead Print*. Published online 2019. http://jtd.amegroups.com/article/view/32687
9. Gilbert DN, Kalil AC, Klompas M, Masur H, Winslow DL. IDSA POSITION STATEMENT: Why IDSA Did Not Endorse the Surviving Sepsis Campaign Guidelines. *Clin Infect Dis*. 2017;45:486. doi:10.1093/cid/cix997
10. Spiegel R, Farkas JD, Rola P, et al. The 2018 Surviving Sepsis Campaign's Treatment Bundle: When Guidelines Outpace the Evidence Supporting Their Use. *Ann Emerg Med*. 2019;73(4):356-358. doi:10.1016/j.annemergmed.2018.06.046
11. Rothrock SG, Cassidy DD, Barneck M, et al. Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. *Ann Emerg Med*. 2020;0(0). doi:10.1016/j.annemergmed.2020.04.042
12. Minderhoud TC, Spruyt C, Huisman S, Oskam E, Schuit SCE, Levin MD. Microbiological outcomes and antibiotic overuse in Emergency Department patients with suspected sepsis. *Neth J Med*. 2017;75(5):196-203. Accessed September 12, 2019. http://www.ncbi.nlm.nih.gov/pubmed/28653945
13. Klein Klouwenberg PMC, Cremer OL, van Vught LA, et al. Likelihood of infection in patients with presumed sepsis at the time of intensive care unit admission: A cohort study. *Crit Care*. 2015;19(1). doi:10.1186/s13054-015-1035-1
14. Alam N, Oskam E, Stassen PM, et al. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. *Lancet Respir Med*. 2018;6(1):40-50. doi:10.1016/S2213-2600(17)30469-1
15. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. *Comput Biol Med*. Published online October 2019:103488. doi:10.1016/j.compbiomed.2019.103488
16. Alam N, Van De Ven PM, Oskam E, et al. Study protocol for a Multi-centre, Investigator-initiated, randomized controlled trial to compare the effects of prehospital antibiotic treatment for sepsis patients with usual care after training Emergency Medical Services (EMS) Personnel in Early Recognition (- the Prehospital ANTibiotics Against Sepsis (PHANTASi) trial. *Acute Med*. 2016;15(4):176-184.
17. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med*. 2003;31(4):1250-1256. doi:10.1097/01.CCM.0000050454.01978.3B

18. R: The R Project for Statistical Computing. Accessed May 25, 2020. https://www.r-project.org/

19. Self-Service Data Analytics Platform | Alteryx. Accessed June 11, 2020. https://www.alteryx.com/

20. Olsen CH. Review of the Use of Statistics in Infection and Immunity. *Infect Immun*. 2003;71(12):6689-6692. doi:10.1128/IAI.71.12.6689-6692.2003

21. Lee PH. Is a cutoff of 10% appropriate for the change-in-estimate criterion of confounder identification? *J Epidemiol*. 2014;24(2):161-167. doi:10.2188/jea.JE20130062

22. Twisk JWR. *Inleiding in de toegepaste biostatistiek*. Vol Vierde (on. Bohn Stafleu van Loghum; 2016. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1638989&site=ehost-live

23. Lee PH. Should we adjust for a confounder if empirical and theoretical criteria yield contradictory results? A simulation study. *Sci Rep*. 2014;4(1):1-14. doi:10.1038/srep06085

24. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med*. 2014;42(8):1749-1755. doi:10.1097/CCM.0000000000000330

25. Liu VX, Fielding-Singh V, Greene JD, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. *Am J Respir Crit Care Med*. 2017;196(7):856-863. doi:10.1164/rccm.201609-1848OC

26. Seymour CW, Gesten F, Prescott HC, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *N Engl J Med*. 2017;376(23):2235-2244. doi:10.1056/NEJMoa1703058

27. Juneja D. Severe sepsis and septic shock in the elderly: An overview. *World J Crit Care Med*. 2012;1(1):23. doi:10.5492/wjccm.v1.i1.23

28. Barbash I, Davis B, Seymour CW, Kahn JM. *D94 SEPSIS BIOLOGY AND HEALTH CARE DELIVERY / Mini Symposium Interaction Between Severity of Illness, Diagnostic Uncertainty and Antibiotic Timing in Patients with Suspected Infection*. Accessed September 10, 2020. www.atsjournals.org

29. Sterling SA, Miller WR, Pryor J, Puskarich MA, Jones AE. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-Analysis. *Crit Care Med*. 2015;43(9):1907-1915. doi:10.1097/CCM.0000000000001142

30. Singer M. Antibiotics for sepsis: Does each hour really count, or is it incestuous amplification? *Am J Respir Crit Care Med*. 2017;196(7):800-802. doi:10.1164/rccm.201703-0621ED

31. Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis users' guides to the medical literature. *JAMA - J Am Med Assoc*. 2014;311(4):405-411. doi:10.1001/jama.2013.285063

# SUPPLEMENTARY APPENDIX

**e-Table 1.** Exploratory K-Centroids Diagnostic Data Mining Trials

| Analysis Trial Number | K-Centroids Method | Min/Max Cluster Parameters | Number of Traits Evaluated | Traits Assessed | Number of Clusters for Partitioning (Based on Preliminary Diagnostics Assessment) |
|---|---|---|---|---|---|
| 1 | K-means | 2/8 | 6 | Sex; Age; Heart Rate (Ambulance); Respiratory Rate (Ambulance); Temperature (Ambulance); Blood Oxygen Saturation (Ambulance) | 4 |
| 2 | K-medians | 2/8 | 6 | Heart Rate (Ambulance); Systolic BP (Ambulance); Diastolic BP (Ambulance); Respiratory Rate (Ambulance); Temperature (Ambulance); Blood Oxygen Saturation (Ambulance) | 3 |
| 3 | K-medians | 2/8 | 2 | Sex; Age | 8 |
| 4 | K-means | 2/8 | 2 | Heart Rate (Ambulance); Temperature (Ambulance) | 5 |
| 5 | K-means | 2/8 | 2 | Age; Temperature (Ambulance) | 3 |
| 6 | K-means | 2/8 | 3 | Age; Heart Rate (Ambulance); Temperature (Ambulance) | 2 |
| 7 | K-means | 2/8 | 2 | Age; Temperature (ED) | 4 |
| 8 | K-means | 2/8 | 2 | Heart Rate (ED); Temperature (ED) | 3 |
| 9 | K-means | 2/8 | 6 | Heart Rate (ED); Systolic BP (ED); Diastolic BP (ED); Respiratory Rate (ED); Temperature (ED); Blood Oxygen Saturation (ED) | 2 |
| 10 | Neural Gas | 2/10 | 3 | Age; Heart Rate (ED); Temperature (ED) | 2 |
| 11 | K-means | 2/8 | 13 | Sex; Age; Heart Rate (Ambulance); Systolic BP (Ambulance); Diastolic BP (Ambulance); Respiratory Rate (Ambulance); Temperature (Ambulance); Blood Oxygen Saturation (Ambulance); Delta Heart Rate (Ambulance > ED); Delta Systolic BP (Ambulance > ED); Delta Diastolic BP (Ambulance > ED); Delta Respiratory Rate (Ambulance > ED); Delta Blood Oxygen Sat (Ambulance > ED) | 2 |

**e-Table 1.** (continued)

| Analysis Trial Number | K-Centroids Method | Min/Max Cluster Parameters | Number of Traits Evaluated | Traits Assessed | Number of Clusters for Partitioning (Based on Preliminary Diagnostics Assessment) |
|---|---|---|---|---|---|
| 12 | K-means | 2/8 | 3 | Age; Respiratory Rate (Ambulance); Delta Respiratory Rate (Ambulance > ED) | 2 |
| 13 | K-means | 2/8 | 2 | Heart Rate (Ambulance); Delta Heart Rate (Ambulance > ED) | |
| 14 | K-means | 2/8 | 4 | Respiratory Rate (Ambulance); Delta Respiratory Rate (Ambulance > ED); Heart Rate (Ambulance); Delta Heart Rate (Ambulance > ED) | 2 |
| 15 | K-medians | 2/8 | 4 | Respiratory Rate (Ambulance); Delta Respiratory Rate (Ambulance > ED); Heart Rate (Ambulance); Delta Heart Rate (Ambulance > ED) | 2 |
| 16 | K-means | 2/8 | 3 | C-reactive Protein (ED Lab); Leucocytes (ED Lab); Creatinine (ED Lab) | 4 |
| 17 | K-means | 2/8 | 2 | Age; Temperature (ED) | 3 |
| 18 | K-means | 2/8 | 3 | Delta Heart Rate (Ambulance > ED); Delta Respiratory Rate (Ambulance > ED); Delta Blood Oxygen Sat (Ambulance > ED) | 2 |
| 19 | K-means | 2/8 | 2 | Heart Rate (ED); Temperature (ED) | 2 |
| 20 | K-means | 2/8 | 5 | Age; Heart Rate (Ambulance); C-reactive Protein (ED Lab); Leucocytes (ED Lab); Creatinine (ED Lab) | 4 |
| 21 | K-means | 2/8 | 5 | Age; Heart Rate (Ambulance); C-reactive Protein (ED Lab); Leucocytes (ED Lab); Creatinine (ED Lab) | 4 |
| 22 | K-means | 2/8 | 3 | C-reactive Protein (ED Lab); Leucocytes (ED Lab); Creatinine (ED Lab) | 4 |

**e-Table 2.** P-values of the interaction term between age and intervention for different cut-off values for age in the full model

| Cut-off (years) | P-value interaction term | Odds ratio interaction term | Confidence interval | Number of young patients | Number of elderly patients |
|---|---|---|---|---|---|
| 70 | 0.255 | 1.58 | 0.72-3.47 | 887 | 1730 |
| 71 | 0.315 | 1.49 | 0.68-3.25 | 936 | 1681 |
| 72 | 0.222 | 1.57 | 0.76-3.27 | 992 | 1625 |
| 73 | 0.166 | 1.65 | 0.81-3.38 | 1073 | 1544 |
| 74 | 0.130 | 1.72 | 0.85-3.47 | 1132 | 1485 |
| 75 | 0.057 | 1.96 | 0.98-3.94 | 1202 | 1415 |
| 76 | 0.025 | 2.17 | 1.11-4.30 | 1296 | 1321 |
| 77 | 0.016 | 2.24 | 1.17-4.34 | 1388 | 1229 |
| 78 | 0.054 | 1.88 | 0.99-3.60 | 1481 | 1136 |
| 79 | 0.111 | 1.67 | 0.89-3.17 | 1583 | 1034 |
| 80 | 0.060 | 1.84 | 0.98-3.47 | 1666 | 951 |
| 81 | 0.171 | 1.56 | 0.93-2.94 | 1767 | 850 |
| 82 | 0.035 | 2.00 | 1.05-3.83 | 1852 | 765 |
| 83 | 0.041 | 1.98 | 1.03-3.83 | 1932 | 685 |
| 84 | 0.205 | 1.54 | 0.79-3.02 | 2015 | 602 |
| 85 | 0.135 | 1.71 | 0.85-3.49 | 2110 | 507 |

# PART IV

## EPILOGUE

# CHAPTER 11

## INTRODUCING ARTIFICIAL INTELLIGENCE TRAINING IN MEDICAL EDUCATION

Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P

*JMIR medical education 2019*
*doi:10.2196/160487;103488*

# ABSTRACT

## Background

Health care is evolving and with it the need to reform medical education. As the practice of medicine enters the age of artificial intelligence (AI), the use of data to improve clinical decision making will grow, pushing the need for skillful medicine-machine interaction. As the rate of medical knowledge grows, technologies such as AI are needed to enable health care professionals to effectively use this knowledge to practice medicine. Medical professionals need to be adequately trained in this new technology, its advantages to improve cost, quality, and access to health care, and its shortfalls such as transparency and liability. AI needs to be seamlessly integrated across different aspects of the curriculum.

Therefore, the aims of this study were to:

1. Understand the impact of AI on health care
2. Address the state of medical education at present
3. Recommended a framework on how to evolve the medical education curriculum to include AI.

## Methods

A systematic literature search was performed searching the electronic database PubMed from inception to Jun 2019 with search terms related to AI, medical education, clinical curriculum, and continuing medical education. Studied the existing medical curriculum in the United States including the Medical College Admission Test (MCAT), United States Medical Licensing Examinations (USMLE), Accreditation Council for Graduate Medical Education (ACGME), Continuing Medical Education (CME), along with AI fundamentals, Electronic Health Records (EHR) training, data sciences, and developed a training framework.

## Results

Introduced a training framework starting from modification to MCAT, introduction of high-quality web-based and face to face data sciences and AI fundamental courses during the core phase of medical education to introductory and refresher courses for attending physicians to extensive training in specific disciplines like radiology, pathology, and clinical decision support for residents and specialists.

## Conclusion

Medical professionals need to be adequately trained in AI, its advantages to improve cost, quality, and access to health care, and its shortfalls such as transparency and liability. AI needs to be seamlessly integrated across different aspects of the curriculum. We recommend a framework on how to evolve the medical education curriculum to include knowledge of AI, data sciences, Electronic Health Records fundamentals, and ethics and legal issues concerning AI. Medical schools will need

to include these in their curriculum to train medical students, residents, fellows, and practicing physicians. A staged approach to education the medical student through journey is recommended.

## Keywords

algorithm; artificial intelligence; black box; deep learning; machine learning; medical education; continuing education; data sciences; curriculum

## Highlights

1. Articulated the state of the art in medical education today
2. Blended technology training (AI, EHRs, data sciences) with medical curriculum starting from MCAT through core medical training phase to clinical phase, residency and specialty training and recommended new training per medical education stage.

## TRENDS IN HEALTH CARE

Global health care expenditure has been projected to grow from USD $7.7 trillion in 2017 to USD $10 trillion in 2022 at a rate of 5.4% [1]. This translates into health care being an average of 9% of gross domestic product among developed countries [2] [3]. Some key global trends that have led to this include tax reform and policy changes in the United States (US) that could impact the expansion of health care access and affordability (Affordable Care Act) [4], implications on the United Kingdom's health care spend based on the decision to leave the European Union [5], population growth and rise in wealth in both China and India [6] [7] [8], implementation of socio-economic policy reform for health care in Russia [9], attempts to make universal health care effective in Argentina [10], massive push for electronic health and telemedicine in Africa [11] and the impact of an unprecedented pace of population aging around the world [12].

From clinicians' perspective there are many important trends that are affecting the way they deliver care of which the growth in medical information is alarming. It took 50 years for medical information to double in 1950. In 1980, it took 7 years. In 2010, it was 3.5 years and is now projected to double in 73 days by 2020 [13]. This growth is posing a challenge to health care professionals to both retain and use it effectively to practice medicine.

## RISE OF ARTIFICIAL INTELLIGENCE IN HEALTH CARE

### Artificial Intelligence in Health Care



Artificial Intelligence (AI) is a scientific discipline that focuses on understanding and creating computer algorithms that can perform tasks that are usually characteristics of humans [14]. AI is now gaining momentum in health care. From its early roots in Sir Alan Turing's seminal paper, Computing Machinery and Intelligence [15], where he proposed the question "Can machines think?", AI has come a long way. Examples of advances in AI include natural language processing (NLP) [16], speech recognition [17] [18], virtual agents[19], decision management[20], machine learning[21], deep learning[22], and robotic process automation [23].

Today, AI is being piloted in health care [24] for faster and accurate diagnosis, to augment radiology [25], reduce errors due to human fatigue, decrease medical costs [26], assist and replace dull, repetitive and labor-intensive tasks [27], minimally invasive surgery [28], and reduce mortality rates [29].

### Challenges With Artificial Intelligence

The rise of AI in health care and its integration into routine clinical practice is going to be a challenge. Along with changing the conventional ways physician work, the *black box problem* [30] and liability issues [31] are some of the most anticipated challenges.

#### *Black Box*

Researchers at Mount Sinai Hospital have created a deep learning algorithm that was trained on the data of 700,000 patients. This algorithm was able to predict onset of a disease such as schizophrenia with high accuracy [32]. This is even more impressive considering the fact that this

condition is difficult to diagnose even for experts. The main problem with this algorithm is that there is no way to know how the system created this prediction and what factors were taken into consideration. This phenomenon is called the *black box* phenomenon. It would not be a precedent in medicine, nevertheless it is difficult to trust a system when there is no understanding on how it works. The physician needs to understand the inputs and the algorithm and interpret the AI proposed diagnosis to ensure no errors are made. We also need to understand what the consequences or unintended side effects are of black box medicine, even when good outcomes can be demonstrated against a standard of care.

Finally, many of the AI systems attempt to mimic aspects of human and animal central nervous systems that are, at large, still a black box. In a recent paper, Zador [33] argued that we have much more to learn from animal brains, in order to unravel this phenomenon.

### *Privacy and Control Over Data*

The development of AI algorithms almost as a rule requires data from a large number of patients. Google, for example, is using 46 billion data points collected from 216,221 adults' de-identified data over 11 combined years from 2 hospitals to predict the outcomes of hospitalized patients [34] [35]. This raises many concerns including relating to patient privacy and control. What happens if a patient does not want to participate in a study where their information is used in algorithm development? In the European Union, the Right to be Forgotten would allow personal data to be erased when the patient has withdrawn their consent [36]. In situations where patient data are limited, algorithm developers train the models on synthetic or hypothetical data, with the risk of generating unsafe and incorrect treatment recommendations [37]. Finally, AI systems are also vulnerable to cybersecurity attacks that could cause the algorithm to misclassify medical information [38].



### *Lack of standards for use of AI in patient care and liability*

Another unresolved question related to the use of AI in health care is liability for the predictions of an algorithm. It is unclear who is liable when a patient experiences serious harm because of an inaccurate prediction. One could argue for any of the involved parties: the physician, the hospital, the company that developed the software, the person who developed the software or even the person who delivered the data. Standards for use of AI in health care are still being developed [39] [40]. New standards for clinical care, quality, safety, malpractice, and communication guidelines have to be developed to allow for greater use of AI. A recently launched AI system for autonomous detection of diabetic retinopathy carries medical malpractice and liability insurance [41] [42].

As use of AI and proactive use of tools such as chatbots [43] increases, physicians and patients will need to be aware of strengths and limitations of such technologies and be trained in how to effectively and safely use them [44] [45].

## How can Artificial Intelligence Address Today's Physician Challenges?

With medical information growing at a breakneck speed, physicians are having trouble keeping up. This is leading to information overload and creates pressure to memorize all this content to

pass the United States Medical Licensing Examinations (USMLE) to qualify for residency positions. Physicians today are working longer hours and are also expected to deliver coordinated care [46] [47] in an aging society with complex conditions and comorbidities where health care costs are increasing and regulations are putting an additional burden on administrative processes.

AI could help physicians by amalgamating large amounts of data and complementing their decision-making process to identify diagnosis and recommend treatments. Physicians in turn need the ability to interpret the results and communicate a recommendation to the patient. In addition, AI could have an impact by alleviating the burden from physicians for performing day-to-day tasks [48]. Speech recognition could help with replacing the use of keyboards to enter and retrieve information [49]. Decision management can help with sifting enormous amounts of data and enable the physician to make an informed and meaningful decision [50] [51]. Automation tools can help with managing regulatory requirements such as Protecting Access to Medicare Act (PAMA) and enable physicians to review the appropriate criteria before making a cost decision [52]. Finally, to help with the acute shortage of health care professionals, virtual agents could in the future help with some aspects of patient care and become a trusted source of information for patients [53].

## ARTIFICIAL INTELLIGENCE TRAINING IN MEDICAL EDUCATION

### State of Medical Education Today

Physicians go through extensive periods of training before they can eventually register as specialists. Although medicine has seen major changes over the last decades, medical education is still largely based on traditional curricula [54]. The specific length of training differs between countries, but the core competencies of these curricula are globally similar[55]. After a core phase of preclinical didactics, training is mostly centered around practice-based learning [56]. Medical education is often based on 6 domains: patient care, medical knowledge, interpersonal and communication skills, practice-based learning and improvement, professionalism, and systems-based practice [57]. These fields were introduced by the Accreditation Council for Graduating Medical Education (ACGME). A large part of medical training focuses on consuming as much information as possible and learning how to apply this knowledge to patient care. This process is still largely memorization based [58]. Less time is spent on familiarizing medical students or residents with new technologies such as AI, mobile health care applications and telemedicine [56] [57] [58]. In the United States, USMLE does not test on these subjects [59]. However, change seems inevitable since the 2018 annual meeting of the American Medical Association (AMA) saw the adoption of AMA's first policy on augmented intelligence, encouraging research into how AI should be addressed in medical education [60]. In Table 1, several initiatives for incorporating AI in medical education are shown, as presented by the AMA [61].

Another important technology-related aspect that is often overlooked in medical training is working with electronic health records (EHRs). EHRs have many benefits, such as improved patient safety, but also assist the implementation of AI in health care. AI algorithms use information from EHR, and therefore the knowledge on how to input unbiased data into the EHR is essential. Otherwise, the AI algorithm will likely be biased as well [62]. At present, training on use of EHR's

**Table 1.** Initiatives for AI in medical education [61]

| Institution | Project |
|---|---|
| Duke Institute for Health Innovation (DIHI) | Medical students work together with data experts to develop care-enhanced technologies made for physicians. |
| University of Florida | Radiology residents work with a technology-based company to develop computer-aided detection for mammography's. |
| Carle Illinois College of Medicine | Offers a course by a scientist, clinical scientist and engineer to learn about new technologies. |
| Sharon Lund Medical Intelligence and Innovation Institute (MI3) | Organizes a summer course on all new technologies in health care, open to medical students. |
| Stanford University Center for Artificial Intelligence in Medicine and Imaging | Involves graduate and post-graduate students in solving heath care problems with the use of machine learning. |
| University of Virginia Center for Engineering in Medicine | Involves medical students in the engineering labs to create innovative ideas in health care. |

for medical students and physicians is not commonly incorporated in the medical curriculum [63], resulting in the medical professional using the EHR as a replacement to capture information on paper without understanding the true potential of this technology [64]. Training on the use of EHR's usually consists of ad hoc brief introductory courses that just teach the basic skills to use the hospital's system in practice. Quality of data and concerns on the impact of the computer on the patient-physician relationship are rarely addressed [63] and the USMLE does not test on these subjects either [59].



## How Clinical Practice is Changing

With the rapid digitization of health care, EHRs facilitate new ways to acquire and process valuable information that can be used to make an informed decision [65]. These advances and transitioning from an information age to the age of AI [58] change clinical practice and patient outcomes for the better. Physicians of the future will have to add to the armory of their skills and competencies, the ability to manage data, supervise AI tools and use AI applications to make informed decisions.

Physicians will have a crucial role in deciding which of these tools is best for their patients. In turn, this will likely change the physician-patient relationship [66]. When information processing is done mainly by computers, this highlights one of the major benefits of AI in medicine: it allows the physician to focus more on caring for and communicating with patients [67]. Finally, in the age of AI, "the physician should combine narrative, mechanistic and mathematical thinking in their training and consider the biopsycho-social model of the disease with the patient at its center". "Computers will never substitute for self-reflective medical expert who is aware of the strengths and limitations of human beings and of an environment characterized by information overload" [68] [69].

## What Will Be Asked From Physicians in the Future?

Future physicians will need a broad range of skills to adequately use AI in clinical practice. Besides understanding the principles of medicine, physicians will also need to acquire satisfactory knowledge of mathematical concepts, AI fundamentals, data science and corresponding ethical and legal issues. These skills will help them to use data from a broad array of sources, supervise AI tools and recognize cases where algorithms might not be as accurate as expected [70]. Furthermore, communication and leadership skills as well as emotional intelligence will be more important than ever as AI-based systems will not be able to consider all the physical and emotional states of the patient [58]. These traits are hard to master for computers and will characterize a great physician in the age of AI.

## Practical Considerations

Some of the time that was originally spent on memorizing medical information will now have to be devoted to other skills. This will have a major impact on the way students and residents will experience their training. The system has to change in such a way that competence will no longer be judged based on factual knowledge but rather on communication skills, emotional intelligence and knowledge on how to use computers.

With an overfull curriculum, there is limited interest in adopting new topics [71], although a 2016 survey by AMA shows that 85% of physicians perceive benefits from new digital tools [61]. The integration of AI-oriented education into the medical curriculum will take time as the technology evolves. A new infrastructure for learning has to be introduced, and new educators from disciplines such as computer sciences, mathematics, ethnography and economics will need to be hired. At the moment, these subjects are not even covered by the core competencies of ACGME, but these competencies "are robust enough to adapt to changing knowledge" [72].

To achieve a change in curriculum, many political and bureaucratic hurdles have to be overcome. *Educational systems, program structures and objectives* have to change in order to create new learning outcomes [73]. A change can only be implemented when large amount of evidence is generated. We have not reached that stage of implementing changes for AI. Furthermore, many other fields within medicine argue that they have not received the attention they deserve [74] [75]. AI needs to prove its benefits and also justify that it is an important topic for medical curriculum over other important subjects that lack adequate medical training at present.

However, one of the most compelling arguments for the implementation of AI training in medical education is that this training will augment existing curriculum rather than replace existing coursework. When students are trained to use AI tools, focus should shift from acquiring basic knowledge on how to use the tool to a basic understanding of the underlying principles. This will enable the students to use this fundamental knowledge when current tools get outdated and new tools are introduced.

Another practical problem is that traditional medical training revolves mainly around the interactions between an attending physician and the residents or medical students. When AI is increasingly introduced into clinical practice, this could be problematic. Many senior physicians

have little to no experience with AI. AI training could be delivered via Continuing Medical Education (CME) programs and might need to be also taught by educators from outside the medical community. For example, a 2-credit CME course on Artificial Intelligence and the Future of Clinical Practice is delivered by a computational biologist and business economists [76].

## RECOMMENDATIONS

### Framework

The traditional medical curriculum, which is mostly memorization based, must follow the transition from the information age to the age of AI. Future physicians have to be taught *competence in the effective integration and utilization of information from a growing array of sources* [58]. To embed this knowledge into medicine, it is of the essence to start introducing these concepts from the beginning of training. In many countries, a Medical College Admission Test (MCAT) has to be taken to be admitted into medical school. The current United States MCAT exam, for example, focuses on biology, chemistry, physics, psychology, sociology and reasoning [81]. These exams could start testing on mathematical concepts such as basis of linear algebra and calculus. These concepts are vital to the elementary understanding of AI and will set the tone for the rest of the curriculum.

In the core phase of preclinical didactics, time should be devoted to working with health data curation and quality [82], provenance [83], integration [84] and governance, working with EHR's [85], AI fundamentals, and ethics and legal issues with AI [86] [87]. Course work in critical appraisal and statistical interpretation of AI and robotic technologies is also important [88]. First, these subjects could be taught in self-contained courses, to teach about the fundamentals of these subjects that can be used even after current applications become outdated [89]. These self-contained courses could potentially replace and augment courses on medical informatics and statistics in the current



**Table 2.** List of Continuing Medical Education programs on artificial intelligence in health care.

| Program | Faculty; *Organization* | Number of Continuing Medical Education credits |
|---|---|---|
| Artificial Intelligence and the Future of Clinical Practice [76] | Computational biologist, Business economist; *Massachusetts Medical Society* | 2.0 |
| Intro to AI and Machine Learning: Why All the Buzz [77] | Medical Informatics, Radiology; *The Radiological Society of North America* | 1.0 |
| Current Applications and Future of Cardiology [78] | Health care Technologists, Bioinformatics, Cardiology; *Mayo Clinic* | 10.0 |
| Artificial Intelligence and Machine Learning: Application in the Care of Children [79] | Pediatric Medicine; *University of Pittsburgh School of Medicine* | 1.0 |
| Artificial Intelligence in Health care: The Hope, The Hype, The Promise, The Peril [80] | Medical Informatics, Business Administration; *Stanford University School of Medicine* | 6.0 |

curriculum. Second, they should also recur in clinical courses to familiarize students with the clinical applications of AI and work with EHR's in diverse settings [89]. An approach to introducing AI could be to incorporate this technology during courses such as Evidence Based Medicine [90]. As the student is taught to appraise evidence through databases such as PubMed or diagnostic tests or systematic reviews, this process could be augmented by applying concepts from data sciences, applying AI technologies such as NLP and analyzing scenarios to test them on questions of ethics and liability [91]. In addition, the students should also be trained in the fundamentals of computer and software engineering to understand the semantics behind real-world AI applications. For example, basics of hardware and software development and user experience design may also be valuable.

During clinical rotations and residency, focus should shift towards relevant applications of AI in practice. With advancements in digital biomarkers [92] and digital therapeutics [93], students should also be trained in these technologies as they rely on AI. They have the potential to enable large-scale diagnostics and treatments in in-home environments in the near future [94]. At the end of training, the USMLE should include a substantial number of questions on data science and AI fundamentals in their final exams. Attendance of conferences on health care AI could be incentivized, so that health care professionals stay up-to-date with the latest developments. For attending physicians, extensive courses on AI and data science should be part of CME. See Table 2 for more details.

AI skills must also be balanced with non-analytics and person-centered aspects of medicine to develop a more rounded doctor of the future. Other skills such as *communications, empathy, shared decision making, leadership, team building and creativity are all skills that will continue to gain importance for physicians*. At the Dell Medical School at the University of Texas, Austin, the curriculum in basic sciences has been reduced in duration to accommodate training in soft skills such as leadership, creativity, and communication [95].



To enable clinicians to think innovatively and create technology-enabled care models, multi-disciplinary training is needed in implementation science, operations and clinical informatics. The Stanford medical school has created such a program to train clinician-innovators for the digital future by introducing a human-centered design approach to graduate medical education [96]. At the Health care Transformation Laboratory at Massachusetts General Hospital in Boston, a 1-year fellowship is offered in health care innovation exposing resident trainees to topics in data sciences, machine learning, health care operations, services, design thinking, intellectual property, and entrepreneurship [97]. These projects are new developments and are the first steps taken in order to introduce AI in medical education.

## First steps

As not all of these interventions can be introduced simultaneously, we suggest a few first steps that will lay the foundation for the upcoming years. We suggest to start off by introducing questions on mathematical concepts into the MCAT similar to the mathematics section in the Graduate Record Examination. High quality web-based courses on data sciences and AI fundamentals should be freely offered in the core phase of medical education. This might lead to students focusing on applications of these subjects more naturally in following years of training.

For residents and medical students who have already finished this phase of training, courses on the fundamental subjects should be available and mandatory throughout the remaining part of their medical education. For students interested in creating new technology-enabled care models, dedicated training in health care innovation during a gap year during the clinical years or after residency should be encouraged. For attending physicians, introductory courses and refresher courses should also be made available. Extensive training is especially necessary for this group so that they can partly take back the task of educating medical students and residents on these subjects in the future. Table 3 lists suggested content that can be added to the various phases of medical education. Table 4 lists a small subset of rapidly evolving AI in health care conferences that physicians and trainees can attend to learn more about this technology and its applications in health care.

**Table 3.** Recommendations per stage of medical education.

| Medical Education Stage | Recommendations | Suggested Content |
|---|---|---|
| MCAT[a] | Introduce questions on linear algebra (vectors, linear transformations, matrix, solutions for linear systems), calculus (limits. Differential calculus, integral calculus), probability (joint, conditional, distribution) | Education Testing Services'(ETS) Graduate Record Examination (GRE) mathematics test [98] |
| Medical School – Core Phase | Working with medical data sets (curation, quality, provenance, integration, governance), EHRs[b], AI[c] fundamentals, Ethics and Legal | Data sets<br>• HealthData.gov [99]<br>• Public datasets in health care [100]<br>• University of California San Francisco Data Resources [101]<br>AI fundamentals<br>• AI 101 course from MIT[d] [102]<br>Ethics, Law<br>• Teaching AI, Ethics, Law and Policy [103]<br>• AI Law [104]<br>EHR Training [105] |
| Medical School – Clinical Phase | Familiarize with AI based clinical applications, Expand knowledge beyond basic principles of data/AI | Clinical Utility<br>• Overview of Clinical applications of AI [106]<br>• AI for Health and Health Care (US Department of Health and Human Services) [107]<br>Center for AI in Medicine and Imaging [108]<br>AI in Health care Accelerated Program [109] |
| USMLE[e] | Introduce questions on data sciences, AI, working with EHRs | Data Science Courses [110] [111] [112] |

**Table 3.** Recommendations per stage of medical education.

| Medical Education Stage | Recommendations | Suggested Content |
|---|---|---|
| Residents | Detailed knowledge on clinical applications, Attend conference in health care AI | Table 4 |
| Specialist | Stay up to date on Data/AI through CME[f] credits, Attend conference in health care AI | Table 2, Table 4 |

[a]MCAT: Medical College Admission Test.
[b]EHRs: electronic health records.
[c]AI: artificial intelligence.
[d]MIT: Massachusetts Institute of Technology
[e]USMLE: United States Medical Licensing Examinations
[f]CME: Continuing Medical Education.

**Table 4.** List of Artificial Intelligence in Health care conferences

| Name of Conference | Topics |
|---|---|
| Ai4 Artificial Intelligence Health care Conference [113] | Exploring top use cases of AI and Machine Learning (ML) in health care |
| AI in Health care [114] | Business value outcomes of AI, Experience in clinical care and hospital operations |
| Machine Learning and AI forum (Health care Information and Management Systems Society - HIMSS) [115] | Data, Analytics, Real-world applications of ML and AI |
| AI in Health care @ JP Morgan Health care Conference [116] | AI applications - drug discovery, secure data exchange, insurer coordination, medical imaging, risk prediction, at-home patient care, and medical billing |
| Radiology in the age of AI [117] | AI in medical imaging |
| American Medical Informatics Association (AMIA) Clinical Informatics Conference [118] | AI in medical informatics |
| Association for the Advancement of Artificial Intelligence (AAAI) [119] | "Increase public understanding of AI, improve the teaching and training of AI practitioners, and provide guidance for research planners and funders concerning the importance and potential of current AI developments and future directions" |

## CONCLUSIONS

Physicians and machines working in combination have the greatest potential to improve clinical decision-making and patient health outcomes [120]. AI can curate and process more data such as medical records, genetic reports, pharmacy notes, and environment data and in turn retain, access, and analyze more medical information. However, it cannot replace the art of caring. As AI and its application become mainstream in health care, medical students, residents, fellows and practicing

physicians need to have knowledge of AI, data sciences, EHR fundamentals, and ethics and legal issues concerning AI. Medical schools will need to include them as part of the curriculum. A staged approach to educating the medical student through their journey is recommended.

AI will enable faster and accurate diagnosis, augment radiology, reduce errors due to human fatigue, decrease medical costs, assist and replace dull, repetitive and labor-intensive tasks, minimally invasive surgery, and reduce mortality rates.

With the global health care expenditure projected to reach US $10 trillion by 2022, AI has the invaluable potential to advance the quadruple aim in health care – enhance the patient experience, improve population health, reduce costs, and improve the provider experience [121] [122].

## CONFLICTS OF INTEREST

KP has written this paper as part of his PhD studies. He is a Vice President at Roche. There is no conflict of interest with his employment at Roche. None of the rest of the authors declare any conflicts of interest.

## ABBREVIATIONS

AI: artificial intelligence
ACGME: Accreditation Council for Graduating Medical Education
AMA: American Medical Association
CME: Continuing Medical Education
EHR: electronic health record
MCAT: Medical College Admission Test
NLP: natural language processing
USMLE: United States Medical Licensing Examinations

## REFERENCES


1. WHO | Current health expenditure as a percentage of gross domestic product (GDP). WHO 2018.
2. American Health Care: Health Spending and the Federal Budget | Committee for a Responsible Federal Budget n.d. https://www.crfb.org/papers/american-health-care-health-spending-and-federal-budget (accessed August 25, 2019).
3. Health Care Spending in US, Other High-Income Countries n.d. https://www.commonwealthfund.org/publications/journal-article/2018/mar/health-care-spending-united-states-and-other-high-income (accessed August 25, 2019).
4. About the ACA | HHS.gov n.d. https://www.hhs.gov/healthcare/about-the-aca/index.html (accessed July 27, 2019).
5. Fahy N, Hervey T, Greer S, Jarman H, Stuckler D, Galsworthy M, et al. How will Brexit affect health and health services in the UK? Evaluating three possible scenarios. Lancet (London, England) 2017;390:2110–8. doi:10.1016/S0140-6736(17)31926-8.
6. China's healthcare provider market: Riding the waves of reform | Deloitte China | Life Sciences and Health care n.d. https://www2.deloitte.com/cn/en/pages/life-sciences-and-healthcare/articles/china-healthcare-provider-market.html (accessed July 27, 2019).
7. Global Burden of Disease Health Financing Collaborator Network JL, Sadat N, Chang AY, Fullman N, Abbafati C, Acharya P, et al. Trends in future health financing and coverage: future health spending and universal health coverage in 188 countries, 2016-40. Lancet (London, England) 2018;391:1783–98. doi:10.1016/S0140-6736(18)30697-4.
8. Horton R. Offline: The new politics of health in India. Lancet (London, England) 2018;392:902. doi:10.1016/S0140-6736(18)32211-6.
9. The Crisis of Russian Health Care and Attempts at Reform | RAND n.d. https://www.rand.org/pubs/conf_proceedings/CF124/CF124.chap5.html (accessed July 27, 2019).
10. Rubinstein A, Zerbino MC, Cejas C, López A. Making Universal Health Care Effective in Argentina: A Blueprint for Reform. Heal Syst Reform 2018;4:203–13. doi:10.1080/23288604.2018.1477537.
11. How new tech can propel Africa to the forefront of healthcare | World Economic Forum n.d. https://www.weforum.org/agenda/2018/01/africa-lead-healthcare-fourth-industrial-revolution/ (accessed July 27, 2019).
12. Ageing and health n.d. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed October 14, 2019).
13. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 2011;122:48–58.
14. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med 2019:103488. doi:10.1016/J.COMPBIOMED.2019.103488.
15. Computing Machinery and Intelligence on JSTOR n.d. https://www-jstor-org.iclibezp1.cc.ic.ac.uk/stable/2251299?seq=1#metadata_info_tab_contents (accessed July 28, 2019).
16. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform 2017;73:14–29. doi:10.1016/j.jbi.2017.07.012.
17. Gold B, Morgan N, Ellis D. Speech and Audio Signal Processing. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011. doi:10.1002/9781118142882.
18. Gold B, Morgan N, Ellis DPW (Daniel PW. Speech and audio signal processing : processing and perception of speech and music. Wiley-Blackwell; 2011.
19. Shaked NA. Avatars and virtual agents – relationship interfaces for the elderly. Healthc Technol Lett 2017;4:83–7. doi:10.1049/htl.2017.0009.
20. Artificial intelligence can now explain its own decision making n.d. https://medium.com/datadriveninvestor/artificial-intelligence-can-now-explain-its-own-decision-making-71fe14d2f53f (accessed July 28, 2019).
21. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. RadioGraphics 2017;37:505–15. doi:10.1148/rg.2017160130.

22. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018;19:1236–46. doi:10.1093/bib/bbx044.

23. Alexovič M, Dotsikas Y, Bober P, Sabo J. Achievements in robotic automation of solvent extraction and related approaches for bioanalysis of pharmaceuticals. J Chromatogr B 2018;1092:402–21. doi:10.1016/j.jchromb.2018.06.037.

24. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2:719–31. doi:10.1038/s41551-018-0305-z.

25. Dreyer KJ, Geis JR. When Machines Think: Radiology's Next Frontier. Radiology 2017;285:713–8. doi:10.1148/radiol.2017171183.

26. Guo J, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. Heal Equity 2018;2:174–81. doi:10.1089/heq.2018.0037.

27. 3 ways medical AI can improve workflow for physicians | American Medical Association n.d. https://www.ama-assn.org/practice-management/digital/3-ways-medical-ai-can-improve-workflow-physicians (accessed February 3, 2019).

28. Wang Z, Majewicz Fey A. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. Int J Comput Assist Radiol Surg 2018;13:1959–70. doi:10.1007/s11548-018-1860-1.

29. Weng SF, Vaz L, Qureshi N, Kai J. Prediction of premature all-cause mortality: A prospective general population cohort study comparing machine-learning and standard epidemiological approaches. PLoS One 2019;14:e0214365. doi:10.1371/journal.pone.0214365.

30. Castelvecchi D. Can we open the black box of AI? Nature 2016;538:20–3. doi:10.1038/538020a.

31. Artificial Intelligence and Medical Liability (Part II) | Bill of Health n.d. http://blogs.harvard.edu/billofhealth/2017/02/10/artificial-intelligence-and-medical-liability-part-ii/#more-20718 (accessed December 24, 2017).

32. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016;6:26094. doi:10.1038/srep26094.

33. Zador AM. A critique of pure learning and what artificial neural networks can learn from animal brains. Nat Commun 2019;10:3770. doi:10.1038/s41467-019-11786-6.

34. Google is using 46 billion data points to predict a hospital patient's future — Quartz n.d. https://qz.com/1189730/google-is-using-46-billion-data-points-to-predict-the-medical-outcomes-of-hospital-patients/ (accessed November 1, 2018).

35. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Liu PJ, et al. Scalable and accurate deep learning for electronic health records 2018. doi:10.1038/s41746-018-0029-1.

36. Right to be Forgotten | General Data Protection Regulation (GDPR) n.d. https://gdpr-info.eu/issues/right-to-be-forgotten/ (accessed October 15, 2019).

37. IBM's Watson recommended "unsafe and incorrect" cancer treatments - STAT n.d. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/ (accessed August 10, 2019).

38. Finlayson SG, Chung HW, Kohane IS, Beam AL. Adversarial Attacks Against Medical Deep Learning Systems 2018.

39. Doshi-Velez F. Considerations for the Practical Impact of AI in Healthcare n.d.

40. How should the FDA approach the regulation of AI and machine learning in healthcare? | Rock Health n.d. https://rockhealth.com/how-should-the-fda-approach-the-regulation-of-ai-and-machine-learning-in-healthcare/ (accessed September 1, 2019).

41. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. Npj Digit Med 2018;1:39. doi:10.1038/s41746-018-0040-6.

42. AskScience AMA Series: I'm Michael Abramoff, a physician/scientist, and Principal Investigator of the study that led the FDA to approve the first ever autonomous diagnostic AI, which makes a clinical decision without a human expert. AMA. : askscience n.d. https://www.reddit.com/r/askscience/comments/9d5id2/

askscience_ama_series_im_michael_abramoff_a/ (accessed October 21, 2018).

43. Bates M. Health Care Chatbots Are Here to Help. IEEE Pulse 2019;10:12–4. doi:10.1109/MPULS.2019.2911816.

44. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res 2019;21:e12887. doi:10.2196/12887.

45. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019;64:070674371982897. doi:10.1177/0706743719828977.

46. Weaver SJ, Che XX, Petersen LA, Hysong SJ. Unpacking Care Coordination Through a Multiteam System Lens. Med Care 2018;56:1. doi:10.1097/MLR.0000000000000874.

47. Elliott J, Stolee P, Boscart V, Giangregorio L, Heckman G. Coordinating care for older adults in primary care settings: understanding the current context. BMC Fam Pract 2018;19:137. doi:10.1186/s12875-018-0821-7.

48. AI, teamed with physicians' intelligence, could improve care | American Medical Association n.d. https://www.ama-assn.org/practice-management/digital/ai-teamed-physicians-intelligence-could-improve-care (accessed July 28, 2019).

49. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. J Am Med Inform Assoc 2016;23:e169-79. doi:10.1093/jamia/ocv152.

50. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016;375:1216–9. doi:10.1056/NEJMp1606181.

51. Levenson SA. The health care decision-making process framework. Md Med 2010;11:13–7.

52. Hentel KD, Menard A, Mongan J, Durack JC, Johnson PT, Raja AS, et al. What Physicians and Health Organizations Should Know About Mandated Imaging Appropriate Use Criteria. Ann Intern Med 2019;170:880. doi:10.7326/M19-0287.

53. Virtual Health Care Could Save the U.S. Billions Each Year n.d. https://hbr.org/2018/04/virtual-health-care-could-save-the-u-s-billions-each-year (accessed July 28, 2019).

54. Pfeifer CM. A progressive three-phase innovation to medical education in the United States. Med Educ Online 2018;00. doi:10.1080/10872981.2018.1427988.

55. The postgraduate medical education pathway: an international comparison. n.d.

56. Pfeifer CM. A progressive three-phase innovation to medical education in the United States. Med Educ Online 2018;00. doi:10.1080/10872981.2018.1427988.

57. Nasca TJ, Philibert I, Brigham T, Flynn TC. The Next GME Accreditation System — Rationale and Benefits. N Engl J Med 2012;366:1051–6. doi:10.1056/NEJMsr1200117.

58. Wartman SA, Donald Combs C. Medical education must move from the information age to the age of artificial intelligence. Acad Med 2018;93:1107–9. doi:10.1097/ACM.0000000000002044.

59. USMLE ® Content Outline. 2019.

60. Digital Health Study Physicians' motivations and requirements for adopting digital clinical tools. 2016.

61. Digital Health Study Physicians' motivations and requirements for adopting digital clinical tools. 2016.

62. Medical Association A. Augmented Intelligence in Health Care policy report. n.d.

63. Wald HS, George P, Reis SP, Taylor JS. Electronic Health Record Training in Undergraduate Medical Education. Acad Med 2014;89:380–6. doi:10.1097/acm.0000000000000131.

64. Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform 2016;Suppl 1:S48-61. doi:10.15265/IYS-2016-s006.

65. Celi LA, Davidzon G, Johnson AEW, Komorowski M, Marshall DC, Nair SS, et al. Bridging the health data divide. J Med Internet Res 2016;18:1–6. doi:10.2196/jmir.6400.

66. Johnston SC. Anticipating and Training the Physician of the Future. Acad Med 2018;93:1105–6. doi:10.1097/ACM.0000000000002175.

67. Mintz LJ, Stoller JK. A Systematic Review of Physician Leadership and Emotional Intelligence. J Grad Med Educ 2014;6:21–31. doi:10.4300/JGME-D-13-00012.1.

68. de Leon J. Teaching Medical Students How To Think: Narrative, Mechanistic and Mathematical Thinking. Actas Esp Psiquiatr 2018;46:133–45.

69. de Leon J, De las Cuevas C. The Art of Pharmacotherapy. J Clin Psychopharmacol 2017;37:131–7. doi:10.1097/JCP.0000000000000675.

70. Park SH, Do K, Kim S, Park JH, Lim Y. What should medical students know about artificial intelligence in medicine ? 2019:16–21.

71. Shortliffe EH. Biomedical Informatics in the Education of Physicians. JAMA 2010;304:1227. doi:10.1001/jama.2010.1262.

72. DeJong C, Lucey CR, Dudley RA. Incorporating a New Technology While Doing No Harm, Virtually. JAMA 2015;314:2351. doi:10.1001/jama.2015.13572.

73. Rudhumbu N. The role of academic middle managers in the planning and implementation of curriculum change in private higher education institutions in Botswana 2015.

74. Mishra S. Do we need to change the medical curriculum: regarding the pain of others. Indian Heart J 2015;67:187–91. doi:10.1016/j.ihj.2015.05.015.

75. Mogre V, Stevens FCJ, Aryee PA, Amalba A, Scherpbier AJJA. Why nutrition education is inadequate in the medical curriculum: a qualitative study of students' perspectives on barriers and strategies. BMC Med Educ 2018;18:26. doi:10.1186/s12909-018-1130-5.

76. Massachusetts Medical Society: Artificial Intelligence and the Future of Clinical Practice n.d. http://www.massmed.org/Continuing-Education-and-Events/Online-CME/Courses/Artificial-Intelligence/Artificial-Intelligence-and-the-Future-of-Clinical-Practice/ (accessed September 1, 2019).

77. Intro to AI and Machine Learning: Why All the Buzz? (2018) - On Demand Webinar n.d. http://education.rsna.org/diweb/catalog/item/eid/1008312688 (accessed September 1, 2019).

78. Current Applications and Future of Artificial Intelligence in Cardiology 2019, InterContinental San Francisco, San Francisco, California, USA | eMedEvents n.d. https://www.emedevents.com/c/medical-conferences-2019/current-applications-and-future-of-artificial-intelligence-in-cardiology (accessed September 1, 2019).

79. Artificial Intelligence & Machine Learning: Application in the Care of Children n.d.

80. Artificial Intelligence in Healthcare: The Hope, The Hype, The Promise, The Peril - Stanford Center for Continuing Medical Education - Continuing Medical Education (CME) n.d. https://stanford.cloud-cme.com/default.aspx?P=0&EID=34240 (accessed September 1, 2019).

81. The MCAT ® Essentials for Testing Year 2019. n.d.

82. van Ooijen PMA. Quality and Curation of Medical Images and Data. Artif. Intell. Med. Imaging, Cham: Springer International Publishing; 2019, p. 247–55. doi:10.1007/978-3-319-94878-2_17.

83. Council Post: Four Reasons Data Provenance Is Vital For Analytics And AI n.d. https://www.forbes.com/sites/forbestechcouncil/2019/05/22/four-reasons-data-provenance-is-vital-for-analytics-and-ai/#503e0cdb57d6 (accessed August 9, 2019).

84. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). Methods Inf Med 2018;57:e57–65. doi:10.3414/ME17-02-0022.

85. Wald HS, George P, Reis SP, Taylor JS. Electronic Health Record Training in Undergraduate Medical Education. Acad Med 2014;89:380–6. doi:10.1097/ACM.0000000000000131.

86. Price WN. Regulating Black-Box Medicine 2017.

87. Price WN. Medical Malpractice and Black-Box Medicine 2017.

88. The Topol Review – NHS Health Education England n.d. https://topol.hee.nhs.uk/ (accessed August 9, 2019).

89. Shortliffe EH. Biomedical Informatics in the Education of Physicians. JAMA 2010;304:1227. doi:10.1001/jama.2010.1262.

90. Heather McEwen MM. NEOMED Library: Evidence-based Medicine I Course Guide (Archived): Introduction n.d.

91. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. JAMA 2016;316:2368. doi:10.1001/jama.2016.17217.

92. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. Npj Digit Med 2019;2:14. doi:10.1038/s41746-019-0090-4.

93. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital Therapeutics: An Integral Component of Digital Innovation in Drug Development. Clin Pharmacol Ther 2018;104:72–80. doi:10.1002/cpt.1036.

94. Alami H, Gagnon M-P, Fortin J-P. Digital health and the challenge of health systems transformation. MHealth 2017;3:31–31. doi:10.21037/mhealth.2017.07.02.

95. Johnston SC. Anticipating and Training the Physician of the Future. Acad Med 2018;93:1105–6. doi:10.1097/ACM.0000000000002175.

96. Carter J, Bababekov YJ, Majmudar MD. Training for our digital future: a human-centered design approach to graduate medical education for aspiring clinician-innovators. Npj Digit Med 2018;1:26. doi:10.1038/s41746-018-0034-4.

97. Healthcare Transformation Lab's Impact Report 2018 | Healthcare Transformation Lab at Massachusetts General Hospital n.d. http://healthcaretransformation.org/2019/03/05/htl-impact-report-2018/ (accessed August 9, 2019).

98. GRE Mathematics Subject Test (For Test Takers) n.d. https://www.ets.org/gre/subject/about/content/mathematics (accessed September 1, 2019).

99. About | HealthData.gov n.d. https://healthdata.gov/node/1 (accessed September 1, 2019).

100. Public Datasets: Descriptions | sgim.org n.d. https://www.sgim.org/communities/research/dataset-compendium/public-datasets-description# (accessed September 1, 2019).

101. Home | UCSF Data Resources n.d. https://data.ucsf.edu/ (accessed September 1, 2019).

102. Artificial Intelligence | Electrical Engineering and Computer Science | MIT OpenCourseWare n.d. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/ (accessed September 1, 2019).

103. Wilk A. Teaching AI, Ethics, Law and Policy 2019.

104. Artificial Intelligence (AI) | Duke University School of Law n.d. https://law.duke.edu/dclt/ai/ (accessed September 1, 2019).

105. Training Courses | Electronic Health Record (EHR) n.d. https://www.ihs.gov/ehr/training/ (accessed September 1, 2019).

106. An Overview of Clinical Applications of Artificial Intelligence | CADTH.ca n.d. https://www.cadth.ca/dv/ieht/overview-clinical-applications-artificial-intelligence (accessed September 1, 2019).

107. JASON: Artificial Intelligence for Health Care – Federation Of American Scientists n.d. https://fas.org/blogs/secrecy/2018/02/ai-health-care/ (accessed September 1, 2019).

108. Center for Artificial Intelligence in Medicine & Imaging | n.d. https://aimi.stanford.edu/ (accessed September 1, 2019).

109. Program Info – Artificial Intelligence in Healthcare Accelerated Program n.d. http://aihap.mgh.harvard.edu/program-info/ (accessed September 1, 2019).

110. The Data Science Course 2019: Complete Data Science Bootcamp | Udemy n.d. https://www.udemy.com/the-data-science-course-complete-data-science-bootcamp/ (accessed September 1, 2019).

111. Data Science: Online Courses from Harvard, MIT, Microsoft | edX n.d. https://www.edx.org/course/subject/data-science (accessed September 1, 2019).

112. Data Science | Coursera n.d. https://www.coursera.org/specializations/jhu-data-science (accessed September 1, 2019).

113. Ai4 Artificial Intelligence Healthcare Conference 2019 - 2nd Annual - Ai4 n.d. https://ai4.io/healthcare/ (accessed September 1, 2019).

114. AI in Healthcare - AI World Conference & Expo n.d. https://aiworld.com/Healthcare-AI (accessed September 1, 2019).

115. Machine Learning & AI Forum | HIMSS2020 n.d. https://www.himssconference.org/education/specialty-programs/machine-learning-ai-healthcare (accessed September 1, 2019).

116. AI in Healthcare @ JP Morgan Healthcare Conference – MIT n.d. https://www.mitcnc.org/events/ai-in-healthcare/ (accessed September 1, 2019).

117. RSNA Spotlight Course n.d. https://www.rsna.org/spotlight/ai-san-francisco (accessed September 1, 2019).

118. AMIA 2019 Clinical Informatics Conference Topics and Keywords | AMIA n.d. https://www.amia.org/cic2019/topics-keywords (accessed September 1, 2019).

119. Association for the Advancement of Artificial Intelligence n.d. http://aaai.org/ (accessed October 13, 2019).

120. AMA Passes First Policy Recommendations on Augmented Intelligence | American Medical Association n.d. https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence (accessed October 15, 2018).

121. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. Health Aff 2008;27:759–69. doi:10.1377/hlthaff.27.3.759.

122. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12:573–6. doi:10.1370/afm.1713.

# CHAPTER 12

## OVERVIEW AND GENERAL DISCUSSION

Sepsis is an enigmatic syndrome. We struggle to understand and define sepsis, even 5000 years after its symptoms were first described[1]. Physicians currently have a wide arsenal of medical interventions and physiological support devices at their disposal, yet around 11 million patients die of sepsis annually[1,2]. The absence of a gold standard definition has led to heterogeneous study populations, compromising the reproducibility of clinical trial results and the ability to find beneficial therapies[3]. Some even describe the current state as a treatment graveyard[3]. Expert panelists, endorsed by numerous medical societies, try to find their way through the maze of heterogeneous study results and attempt to formulate management strategies that will positively affect most sepsis patients[4]. This is a hard task since the evidence underlying even the most obvious treatment strategies is limited or contradictory. Now that we enter the age of artificial intelligence, advanced analytical techniques may help us better understand sepsis and how to manage it. This thesis aimed to investigate the cornerstones of sepsis management and ways to use data and machine learning to optimize their use. Throughout this thesis, we have discussed potential AI solutions for sepsis. However, the use of AI in everyday clinical practice is still in its infancy. We have seen significant barriers that must be addressed before using these tools for patient care. How to overcome these barriers will form the basis of the following general discussion.

## THE CURRENT STATE OF SEPSIS MANAGEMENT

The international Surviving Sepsis Campaign (SSC) has provided evidence-based guidelines for sepsis care since its establishment in the early 2000s[5]. By providing a standard of care and increasing awareness around sepsis, the SSC aims to reduce morbidity and mortality from sepsis globally. Their recommendations are bundled into groups of similar care processes to be performed within specific timeframes to facilitate implementation. Nevertheless, non-compliance with these recommendations is significant. Individual hospitals have introduced sepsis performance improvement programs to maximize adherence to local or international sepsis protocols. The latest update of the SSC guideline in 2021 recommends that hospitals and health systems use such improvement programs to increase compliance with sepsis guidelines[4]. However, we need more evidence on the optimal structure of such programs and their potential impact. In **chapter 2**, we discussed the literature on using sepsis performance improvement programs. Throughout the literature, they are consistently associated with improved bundle adherence and with lower mortality rates[6]. The most successful programs include combinations of interventions such as screening tools, educational programs, and specialized sepsis response teams. However, it must be emphasized that we need to keep thinking critically about when to deviate from guideline recommendations. The lack of evidence for multiple aspects of effective sepsis management has led to inevitable one-size-fits-all recommendations. These approaches lead us to overuse resources in low-risk patients and perhaps even underuse them in high-risk patients. When we implemented our sepsis performance improvement program in **chapter 3**, we deliberately allowed the physicians to maneuver according to their clinical judgment and find cases where it may be best to deviate from the general recommendations. Nevertheless, deep-rooted beliefs and fears about the benefits of some sepsis-related interventions limit the physician's ability to tailor a management protocol to the specific patient.

One such fear is missing bloodstream infections[7]. Bloodstream infections are associated with high morbidity and mortality rates, and the general recommendation during the work-up of sepsis is to draw blood cultures from all suspected patients[4,8]. Consequently, the yield of blood cultures is low. The percentage of true-positive results in the emergency department or ward is usually below 10%, while there are at least as many false-positive (contaminated) results[9–11]. The liberal use of blood cultures, sometimes without the appropriate indication, puts the patient at risk of serious, though often unnoticed, harms associated with blood culture contamination. Various studies have shown that blood culture contamination is associated with additional resource use, antibiotic therapy, prolonged hospital stays, and in-hospital mortality[11–14]. Diagnostic stewardship interventions to provide swift and personalized suggestions for the diagnostic work-up of sepsis are needed to reduce resource overuse. It may save patients from undergoing painful tests and potentially harmful side effects. In **chapter 5**, we developed a machine learning algorithm to predict blood culture results in the emergency department and conducted a multicenter validation and prospective evaluation. We further implemented the algorithm in our hospital's electronic health record system to study its real-time performance. This machine learning tool can provide physicians with decision support to withhold blood culture analyses in low-risk patients, potentially reducing the number of tests by 30%. In the next step, we will investigate the clinical benefits of using the tool in an RCT, which will be further discussed in the future perspectives section.

Another particularly interesting intervention in sepsis management, where deep-rooted beliefs prevent us from tailoring the treatment effectively, is the early administration of antibiotics. A highly cited retrospective study from 2006 found that every hour in the delay of antibiotic therapy for patients with sepsis and persistent hypotension was associated with decreased survival[15]. Physicians have since been challenged to treat patients with suspected infections as soon as possible while sacrificing diagnostic accuracy. **Chapter 9** reviewed the literature on the benefits of early antibiotics for sepsis. The evidence supporting this practice is based solely on retrospective data, while meta-analyses and the one randomized clinical trial on this subject have failed to show any benefits[16–20]. Some even found harm by aggressive initiation of antimicrobial treatment[21]. Preliminary work on diagnostic uncertainty in sepsis suggests an important interaction between the probability of (bacterial) infection and the benefits or harms of early antibiotics[22]. Those with a high probability of bacterial infection seem to benefit, but those with a low probability may experience harm. In our review, we encouraged physicians to temporarily delay the administration of antibiotics to conduct a rapid, proper assessment of the probability of infectious versus non-infectious causes of the disease in patients without shock. The latest revision of the SSC guidelines includes a similar recommendation, moving from a "one-size-fits-all" to a "few-sizes-fit-most" approach (figure 1). In future iterations, data and machine learning may help standardize and further tailor this "few-sizes-fit-most" approach to personalized medicine.



## THE CURRENT STATE OF CLINICAL SEPSIS RESEARCH

For decades, researchers have investigated new drugs and treatment strategies for sepsis[23]. Unfortunately, this search has yet to yield results. Consequently, the treatment of sepsis still

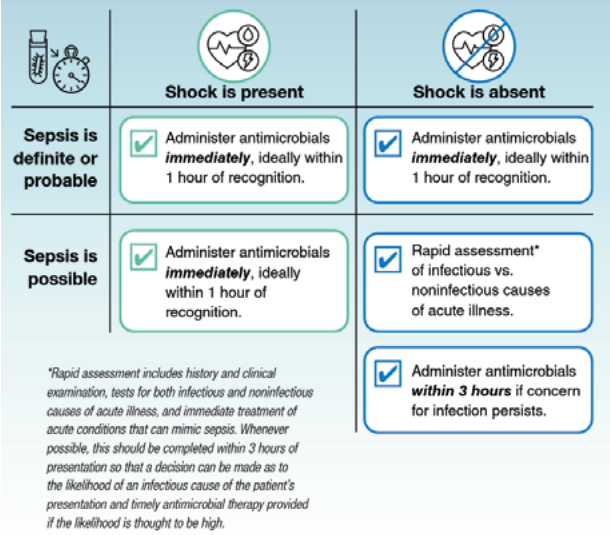



**Figure 1. Recommendations for early antibiotics for sepsis in the 2021 Surviving Sepsis Campaign guidelines (*adapted from the surviving sepsis campaign guidelines[4])**

consists only of broad and nonspecific interventions such as supportive care and infection control[24]. The significant heterogeneity of the sepsis population is a primary reason for the lack of positive trial results for new treatments. Many studies have investigated new therapies among patients with various characteristics and types of infections. Recently, sepsis research has shifted its focus to finding subgroups of patients with similar traits that may respond more similarly to treatments. These subgroups can be found in various ways. Early work aimed to derive sepsis subtypes based on blood genomic endotypes[25]. Others have focused on finding sepsis phenotypes using more widely available data from laboratory tests and vital sign measurements[26]. Over the past years, many sepsis researchers worldwide have tried to find more homogeneous clusters of sepsis patients. However, few have validated these clusters and investigated whether they can be reproduced in other cohorts. The COVID-19 pandemic created a unique opportunity to do just that. The World Health Organization urged researchers to collect data on this new disease following a standardized format provided by the organization. In **chapter 8**, we validated clusters of COVID-19 patients from a French hospital. We had captured the same data points in our Dutch cohort. Although we could replicate the methods and found some broadly overlapping characteristics in the clusters, there were also distinct differences between the French clusters and those we identified. These differences raise the question of whether unsupervised techniques are truly useful in reducing heterogeneity in sepsis cohorts. Experts have argued that useful subtypes must be biologically plausible, treatment-responsive, promptly identifiable, and reproducible[27]. Most unsupervised clustering methods optimally result in three to five clusters, which can still form large and heterogeneous subgroups. Intrinsically, they may be insufficiently distinctive to inform better treatment options, even when

their results are reproducible. Rather than identifying large groups, we may need to tailor treatment suggestions to the individual patient, requiring more supervised and reinforcement learning approaches.

Besides the heterogeneity in the sepsis population, research may also be hampered by the outcome measures it uses. In 2005, the International Sepsis Forum proposed to widen the range of outcome measures investigated in sepsis trials[28]. Mortality benefits are attractive results but capture only the tip of the iceberg. Sepsis survivors experience significant long-term morbidity, which should also be a focus of any intervention trial[28]. Nevertheless, the literature is still dominated by the search for short-term mortality benefits. In **chapter 7**, we learned that for COVID-19, a common and extensive set of outcome measures was successfully adopted worldwide[29]. The advantage of investigating a distinct subset of viral sepsis, which originates from a single pathogen and organ, is that site-specific outcome measures like lung function tests can also be used. Sepsis researchers may adapt trial designs from the COVID-19 field to make stratification per infection type possible and meaningful.

## DATA-DRIVEN APPROACHES TO SEPSIS

Throughout this thesis, we have encountered many opportunities for data and machine learning to help optimize sepsis care. In **chapter 4**, we narratively reviewed some of the clinical applications of AI for sepsis. A large part of that literature focuses on using supervised machine learning methods to detect sepsis early[30–32]. A well-known sepsis detection model is the Targeted Real-Time Early Warning Score (TREWS), which has a reported area under the curve (AUC) of 0.97[33,34]. Supervised methods can also be used to predict the results of diagnostic tests, as we did with our blood culture prediction tool in **chapter 5**. As discussed above, unsupervised machine learning, such as cluster analyses, has already contributed substantially to the recent sepsis literature. By clustering patients with similar traits, we can create endotypes (subgroups with distinct pathobiological mechanisms) or phenotypes (subgroups with distinct observable traits) of sepsis with a higher chance of responding similarly to treatments[25,26]. The degree to which these subtypes can be relevant to clinical practice is still largely unknown[27]. We may need more supervised and reinforcement learning approaches to provide personalized care. In research settings, the latter has already been shown to provide excellent fluid and ventilation strategies for critically ill sepsis patients, increasing the chance of 90-day survival compared to physician policies[35,36].



## WHAT IS NEEDED TO SUCCEED: FUTURE PERSPECTIVES

Before AI can impact patient care, significant barriers must be addressed. In **chapter 6**, we explored those barriers in conversations with physicians. Surprisingly, these open discussions did not focus on some of the well-known issues like data privacy, algorithmic bias, or liability issues[37,38]. Remarkably, the physicians unanimously felt that they should always be liable when poor decisions were made based on the predictions of an AI algorithm, as they would ultimately be making the decision themselves. However, they expressed the need for sufficiently strong evidence that the predictions are accurate and that using them would benefit the patient.

So far, AI tools for healthcare have largely been an academic exercise. While over 50.000 studies of medical AI models can be found through Medline alone, just 224 tools have been approved by the US Food & Drug Administration[39,40]. Similar patterns can be observed for sepsis, with many models developed and few translating to the clinical arena[41,42]. We have only seen the first real-world evaluations of AI-based sepsis detection tools in the past few years. In 2021, the Epic Sepsis Model, a sepsis detection tool by the EHR vendor Epic, was validated[30]. With an area under the curve of 0.63, the performance was much worse than initially reported. In the worst-case scenario, the physicians needed to evaluate 109 patients to detect one sepsis case earlier, putting an additional burden on the healthcare system. Clinical and operational heterogeneity and shifts in patient mix and protocols have caused inevitable performance drifts[43,44]. More recently, the earlier discussed TREWS score for the early detection of sepsis was deployed in five hospitals, and its potential benefit on patient outcomes was evaluated among 6.877 actionable sepsis cases[33]. The study showed that using the alert helped reduce the relative mortality rate by 18.7%. However, there are significant concerns regarding the control group, which may have included many non-septic patients. Despite being one of the most extensive evaluations of a sepsis-related AI tool, the study's observational nature limits our conclusions. Since sepsis alerts can trigger one-size-fits-all protocols, improper use can cause harm through the overuse of antibiotics and the burden put on the physicians when they must evaluate many patients to detect one sepsis case earlier[30,45]. Their implementation should not be taken lightly. We need high-quality evidence to show unequivocally that using this and other AI-based decision-support tools will benefit the patient with sepsis[46]. The authors of the TREWS study rightly stated that large-scale RCTs are exceptionally difficult to carry out[33]. We have accepted this challenge and are currently setting up a multicenter RCT to study the impact of our blood culture prediction tool, described in **chapter 5**, on clinical endpoints. In this trial, we will randomize between doing blood cultures based on the physician's judgment (control) or the algorithm's decision (intervention). Whenever a patient in the intervention group has a probability of less than 5% of a positive blood culture, the test will not be performed or will be canceled when the blood has already been drawn. We will evaluate the impact of this decision-support on patient-related, diagnostics-related, and therapy-related outcomes. We hypothesize that decisions made by following the algorithm's predictions will result in similar (non-inferior) mortality rates and may reduce the number of diagnostic tests (laboratory and microbiology), the duration of antibiotic therapy, and perhaps even the length of stay in the hospital. Setting up an RCT in this space requires significant efforts and documentation, as these tools fall under the Medical Device Regulation. Also, we will need to include over 7500 patients to provide a definitive answer as to whether our tool is safe and beneficial. Nevertheless, we believe this step is necessary to advance the impact of medical AI on patient care.

Finally, the patient is the most important stakeholder and should not be overlooked during the transition to data-driven (sepsis) care. Although AI tools often support the physician, they will ultimately affect the patient. Patients and physicians may have different ethical and moral views about how these technologies should or should not be used. To deliver the best care, patients should be involved in AI development and deployment. Our RCT protocols have consequently been created in partnership with patient representatives.

## CONCLUDING REMARKS

*The complexity of sepsis makes it a difficult condition to diagnose and treat. However, with the help of AI tools, we can gain a better understanding of this deadly syndrome and develop more effective treatments. AI can help us identify patterns in patient data that may be missed by traditional methods, allowing us to better predict outcomes and tailor treatments for individual patients. By leveraging the power of AI, we can make strides towards improving the diagnosis and treatment of sepsis and ultimately saving lives.* To substantiate that claim, I must admit that these concluding remarks were not written by the author of this thesis but by an AI algorithm[47].

# REFERENCES


1. Angus, D. C. & Bindman, A. B. Achieving Diagnostic Excellence for Sepsis. *JAMA* **327**, 117–118 (2022).
2. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
3. Riedemann, N. C., Guo, R. F. & Ward, P. A. The enigma of sepsis. *J. Clin. Invest.* **112**, 460–467 (2003).
4. Evans, L. *et al.* Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* (2021) doi:10.1007/s00134-021-06506-y.
5. Dellinger, R. P. The Future of Sepsis Performance Improvement. *Critical Care Medicine* vol. 43 1787–1789 Preprint at https://doi.org/10.1097/CCM.0000000000001231 (2015).
6. Damiani, E. *et al.* Effect of Performance Improvement Programs on Compliance with Sepsis Bundles and Mortality: A Systematic Review and Meta-Analysis of Observational Studies. (2015) doi:10.1371/journal.pone.0125827.
7. Coburn, B., Morris, A. M., Tomlinson, G. & Detsky, A. S. Does this adult patient with suspected bacteremia require blood cultures? *JAMA - Journal of the American Medical Association* **308**, 502–511 (2012).
8. Goto, M. & Al-Hasan, M. N. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clinical Microbiology and Infection* vol. 19 501–509 Preprint at https://doi.org/10.1111/1469-0691.12195 (2013).
9. Denny, K. J., Sweeny, A., Crilly, J., Maloney, S. & Keijzers, G. Is it time for a culture change? Blood culture collection in the emergency department. *Emergency Medicine Australasia* **30**, 575–577 (2018).
10. Nannan Panday, R. S. *et al.* Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. *PLoS One* **14**, (2019).
11. Klucher, J. M., Davis, K., Lakkad, M., Painter, J. T. & Dare, R. K. Risk factors and clinical outcomes associated with blood culture contamination. *Infect Control Hosp Epidemiol* 1–7 (2021) doi:10.1017/ice.2021.111.
12. Zwang, O. & Albert, R. K. Analysis of strategies to improve cost effectiveness of blood cultures. *J Hosp Med* **1**, 272–276 (2006).
13. Bates, D. W., Goldman, L. & Lee, T. H. Contaminant Blood Cultures and Resource Utilization: The True Consequences of False-Positive Results. *JAMA: The Journal of the American Medical Association* **265**, 365–369 (1991).
14. Dempsey, C., Skoglund, E., Muldrew, K. L. & Garey, K. W. Economic health care costs of blood culture contamination: A systematic review. *American Journal of Infection Control* vol. 47 963–967 Preprint at https://doi.org/10.1016/j.ajic.2018.12.020 (2019).
15. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med* **34**, 1589–1596 (2006).
16. Seymour, C. W. *et al.* Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *N Engl J Med* **376**, 2235–2244 (2017).
17. Liu, V. X. *et al.* The Timing of Early Antibiotics and Hospital Mortality in Sepsis. *Am J Respir Crit Care Med* **196**, 856–863 (2017).
18. Rothrock, S. G. *et al.* Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. *Annals of Emergency Medicine* vol. 0 Preprint at https://doi.org/10.1016/j.annemergmed.2020.04.042 (2020).
19. Sterling, S. A., Miller, R., Pryor, J., Puskarich, M. A. & Jones, A. E. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis HHS Public Access. *Crit Care Med* **43**, 1907–1915 (2015).
20. Alam, N. *et al.* Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. *Lancet Respir Med* **6**, 40–50 (2018).
21. Hranjec, T. *et al.* Aggressive versus conservative initiation of antimicrobial treatment in critically ill surgical patients with suspected intensive-care-unit-acquired infection: A quasi-experimental, before and after observational cohort study. *Lancet Infect Dis* **12**, 774–780 (2012).

22. Barbash, I., Davis, B., Seymour, C. W. & Kahn, J. M. *D94 SEPSIS BIOLOGY AND HEALTH CARE DELIVERY / Mini Symposium Interaction Between Severity of Illness, Diagnostic Uncertainty and Antibiotic Timing in Patients with Suspected Infection*. www.atsjournals.org.

23. Marshall, J. C. Why have clinical trials in sepsis failed? *Trends Mol. Med.* **20**, 195–203 (2014).

24. Vincent, J.-L. Current sepsis therapeutics. *EBioMedicine* **0**, 104318 (2022).

25. Scicluna, B. P. *et al.* Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med* **5**, 816–826 (2017).

26. Seymour, C. W. *et al.* Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* **321**, 2003 (2019).

27. Demerle, K., Angus, D. C. & Seymour, C. W. Precision Medicine for COVID-19: Phenotype Anarchy or Promise Realized? *JAMA - Journal of the American Medical Association* vol. 325 2041–2042 Preprint at https://doi.org/10.1001/jama.2021.5248 (2021).

28. Marshall, J. G. *et al.* Outcome measures for clinical research in sepsis: A report of the 2nd Cambridge Colloquium of the International Sepsis Forum. *Crit Care Med* **33**, 1708–1716 (2005).

29. Marshall, J. C. *et al.* A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases* vol. 20 e192–e197 Preprint at https://doi.org/10.1016/S1473-3099(20)30483-7 (2020).

30. Wong, A. *et al.* External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* **181**, 1065–1070 (2021).

31. Nemati, S. *et al.* An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med* **46**, 547–553 (2018).

32. Desautels, T. *et al.* Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform* **4**, e28 (2016).

33. Adams, R. *et al.* Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nature Medicine 2022 28:7* **28**, 1455–1460 (2022).

34. Henry, K. E. *et al.* Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing. *Nature Medicine 2022 28:7* **28**, 1447–1454 (2022).

35. Peine, A. *et al.* Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. *NPJ Digit Med* **4**, (2021).

36. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* **24**, 1716–1720 (2018).

37. Kogetsu, A., Ogishima, S. & Kato, K. Authentication of patients and participants in health information exchange and consent for medical research: A key step for privacy protection, respect for autonomy, and Trustworthiness. *Front Genet* **9**, 167 (2018).

38. Finlayson, S. G., Chung, H. W., Kohane, I. S. & Beam, A. L. Adversarial Attacks Against Medical Deep Learning Systems. *Science (1979)* **363**, 1287–1289 (2018).

39. Zhang, J. *et al.* An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. *Lancet Digit Health* **4**, e212–e213 (2022).

40. Muehlematter, U. J., Daniore, P. & Vokinger, K. N. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* **3**, e195–e203 (2021).

41. van de Sande, D., van Genderen, M. E., Huiskens, J., Gommers, D. & van Bommel, J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med* 1–11 (2021) doi:10.1007/s00134-021-06446-7.

42. Fleuren, L. M. *et al.* Machine learning in intensive care medicine: ready for take-off? *Intensive Care Medicine* vol. 46 1486–1488 Preprint at https://doi.org/10.1007/s00134-020-06045-y (2020).

43. Davis, S. E. *et al.* A nonparametric updating method to correct clinical prediction model

drift. *Journal of the American Medical Informatics Association* **26**, 1448–1457 (2019).

44. van Calster, B., Wynants, L., Timmerman, D., Steyerberg, E. W. & Collins, G. S. Predictive analytics in health care: how can we know it works? *Journal of the American Medical Informatics Association* vol. 26 1651–1654 Preprint at https://doi.org/10.1093/jamia/ocz130 (2019).

45. Minderhoud, T. C. *et al.* Microbiological outcomes and antibiotic overuse in Emergency Department patients with suspected sepsis. *Neth J Med* **75**, 196–203 (2017).

46. Angus, D. C. Randomized Clinical Trials of Artificial Intelligence. *JAMA* **323**, 1043–1045 (2020).

47. OpenAI. *Input: write some concluding remarks that sepsis is a highly heterogeneous and deadly syndrome and that AI tools can help further unravel the secrets of sepsis. Maximum length: 150. Temperature: 0. F/P penalty: 0.5.* https://openai.com/.

# CHAPTER 13

## SUMMARY

## SUMMARY

This final chapter summarizes the main findings of this thesis. This thesis aimed to investigate the cornerstones of sepsis management, including the diagnostic work-up and antibiotic treatment, and ways to use data and machine learning to optimize their use.

## SEPSIS PERFORMANCE IMPROVEMENT PROGRAMS

We first needed to consider the current baseline of sepsis management to find ways to use data and machine learning to optimize sepsis care. The international Surviving Sepsis Campaign (SSC) has provided evidence-based guidelines for sepsis care since its establishment in the early 2000s[1]. To facilitate guideline implementation, the SSC has bundled its recommendations into groups of similar care processes to be performed within specific timeframes. Nevertheless, non-compliance with these recommendations is significant. In response to low adoption rates, hospitals have created performance improvement programs to improve adherence to local or international sepsis protocols. The latest update of the SSC guideline recommends that hospitals and health systems use such sepsis improvement programs to increase compliance with the guidelines[2]. **Chapter 2** discusses the literature on the use and benefits of sepsis performance improvement programs. We also provide practical insights for clinical implementation. Throughout the literature, sepsis performance improvement programs are consistently associated with improved bundle adherence and with lower mortality rates[3]. Ideally, these programs should integrate screening tools, changes to sepsis care pathways, and educational initiatives to increase awareness about sepsis care. Engaging large multidisciplinary groups of stakeholders, including patients, is essential to implement these programs successfully.

Using what we learned about sepsis performance improvement programs, we implemented such an intervention in our hospital in **chapter 3**. We performed a before-after intervention study in the emergency department of Amsterdam UMC. The intervention consisted of a screening tool, educational meetings, audits and feedback, and a multidisciplinary sepsis response team. The postintervention phase was associated with improvements in most process-related outcomes, such as a shorter time to antibiotics (66 vs. 143 minutes; $p<0.001$), more lactate measurements (72.9% vs. 46.2%; $p<0.001$), and more completed Modified Early Warning Scores (MEWS; 85.0% vs. 62.9%; $p<0.001$) compared with the preintervention phase. However, there were no differences in patient-related outcomes between the preintervention and postintervention phases except for an improved rate of immediate versus delayed intensive care unit admissions (100% immediate vs. 64.3% immediate; $p=0.012$). We conclude that the program stimulated collaborative and timely decision-making and improved protocol adherence while allowing physicians to maneuver according to their clinical judgment. The results may create urgency for a larger (stepped wedge cluster) RCT to fully capture the value of sepsis performance improvement programs.

## THE DIAGNOSTIC WORK-UP OF SEPSIS

Since sepsis can arise from many sources and pathogens, the diagnostic work-up is highly diverse and challenging to capture in a guideline. Artificially intelligent (AI) tools may provide more tailored recommendations by combing through large amounts of data and finding subtle patterns that

humans may overlook. In **chapter 4**, we conducted a narrative review to map all the currently available AI decision support tools for sepsis and explore their potential and pitfalls. By assessing the quality of the included studies using the PROBAST tool, we realized that some cases using AI for sepsis came with significant challenges. AI models which predicted the onset of sepsis often used predictor variables such as blood pressure in the model. This predictor is also part of the sepsis definition, which it tries to predict. Due to this incorporation bias, the performance of these tools is overestimated.

A crucial part of the diagnostic work-up for which few AI models exist is predicting the results of (blood) cultures. Physicians order blood cultures liberally for fear of missing bloodstream infections, which have high morbidity and mortality rates[4,5]. However, ordering blood cultures without the appropriate indication leads to low yields while putting the patient at risk of harm from false positive results. Blood culture contamination (false-positive cultures) is associated with additional microbiological testing, unnecessary use of antibiotics, prolonged hospital stays, and even in-hospital mortality[6]. **Chapter 5** presents how we developed a machine learning algorithm to predict blood culture results in the emergency department, followed by a multicenter validation and prospective evaluation. The area under the curve (AUC) of the model predictions of whether a blood culture would be positive was 0.81 (95%-CI = 0.78–0.83) in Amsterdam UMC location VU university medical center (VUmc) and between 0.75-0.80 in three external cohorts in the Netherlands and the United States. We further implemented the algorithm in the VUmc electronic health record system for real-time evaluation, in which it retained an AUC of 0.76. Using the tool to withhold blood culture testing in low-risk patients, we could potentially reduce the number of blood cultures in the ED by 30%. However, several challenges must first be addressed for successful adoption in practice.

The barriers to and facilitators of clinical AI implementation among healthcare professionals are ill-defined. **Chapter 6** investigates those barriers in a mixed-methods study with physician interviews, focus group discussions, and a nationwide survey. The important constructs (themes) arising from the discussions, such as tension for change, access to knowledge and information, and evidence strength, were then matched to appropriate implementation strategies according to the Expert Recommendations for Implementing Change (ERIC). We conclude that the current tension for change to implement AI needs to be sparked to facilitate sustainable implementation. This can be accomplished using educational meetings and committed local leaders. Healthcare professionals also want trial phases to compare AI recommendations to their judgment to increase confidence before relying on it. Lastly, it is crucially important to them that AI algorithms are tailored to the local context and fit existing workflows. Keeping all these aspects in mind, we further optimized our blood culture prediction algorithm and designed an RCT to investigate its potential benefits on patient outcomes. The trial was further explained in the general discussion.



## OPTIMIZING THE TREATMENT OF SEPSIS

Despite over a hundred RCTs of immunomodulating drugs, the current treatment of sepsis consists only of broad and nonspecific interventions such as supportive care and infection control[7,8].

The significant heterogeneity of the sepsis population is a primary reason for the lack of positive trial results for new treatments. Many studies have investigated new therapies among patients with various characteristics and types of infections. Recently, sepsis research has shifted its focus to finding subgroups of patients with similar features that may respond more similarly to treatments. The coronavirus disease (COVID-19) pandemic unexpectedly reinforced the value of this approach. In **chapter 7**, we describe that new and effective therapies for this distinct subgroup of viral sepsis could be found within months. The chapter details various aspects of COVID-19 research from which sepsis researchers can learn, such as using a broader range of outcome measures, global collaboration with standardized data capture, and studying a more homogeneous population[9].

As COVID-19 research considers only a single pathogen and disease origin, it is essentially different from investigating the wide collection of symptoms and outcomes captured under the "umbrella" term sepsis [8,10]. Even then, there is significant heterogeneity in how patients (hosts) respond to an infection with SARS-CoV-2. In **chapter 8**, we used clustering techniques to find even more homogeneous subgroups within the COVID-19 population. We replicated the methods of a French study to see whether we could find similar sub-phenotypes of COVID-19. Indeed, our clustering analyses on a multicenter Dutch cohort found similar distributions of characteristics to the French study. Sub-phenotype 1 consisted mainly of relatively young female (74.5%) patients with a high prevalence of gastrointestinal complaints but few comorbidities. This sub-phenotype had the most favorable outcome. Sub-phenotype 2 included more male patients (80.4%). They presented with fewer symptoms but had worse outcomes than sub-phenotype 1. Sub-phenotype 3 included older, mostly male patients with various comorbidities. This sub-phenotype was associated with the worst outcomes. Showing that these cluster analyses provide relatively similar subgroups of COVID-19 patients to the French study supports the robustness of the approach. However, there were also distinct differences between the French and Dutch clusters, and we should remain cautious in using these techniques to understand disease heterogeneity.

Following these findings, we would like to apply such strategies to discover groups of sepsis patients who may benefit most from specific interventions. A particularly interesting intervention in this regard is the early administration of antibiotics for sepsis, which may be the closest to a sepsis-specific therapy we have. There is a deep-rooted belief that every hour delay in the administration of antibiotics will decrease the chances of survival, as was first proposed in a highly cited paper from 2006[11]. **Chapter 9** reviewed the literature on the benefits of early antibiotics for sepsis. The evidence supporting this practice is based only on retrospective data, while meta-analyses and the one randomized clinical trial on this subject have failed to show any benefits. Still, the current guidelines often challenge physicians to treat patients with potential sepsis with broad-spectrum antibiotics as soon as possible. In the review, the real challenge was identifying cases where we can safely delay antibiotic treatment to gather additional data to increase or decrease the likelihood of a bacterial infection. Although this general notion is likely to improve sepsis management, we still were unable to find easily identifiable subgroups of patients who would, on a group level, benefit from early antibiotic treatment.

In **chapter 10**, we used cluster analyses to try and identify those subgroups in the study population of the only RCT on this subject, the pre-hospital antibiotics against sepsis (PHANTASi)

trial. Surprisingly, the clustering patterns consistently showed that age was the most important driver of cluster formation. When we subsequently summarized the mortality rates across clusters and intervention groups, there was an interaction between the benefits of early antibiotic treatment and younger age. Further logistic regression modeling to quantify this relationship, adjusted for confounders, confirmed a significant interaction between younger age and the benefits of early antibiotics. When adjusting for this interaction, we could find a significant benefit of early antibiotic treatment (odds-ratio 0.07; 95% CI, 0.01-0.79; *P* = .03) in the complete study population. We can work toward more personalized recommendations for sepsis treatments using these data-driven insights.

## EPILOGUE

The healthcare landscape is evolving, and data and machine learning to support medical decision-making will become increasingly important. In 2020, it was estimated that the total amount of medical information doubled every 73 days[12]. New technologies such as AI can enable the effective use of all this information. This will require skillful medicine-machine interactions, for which healthcare professionals must be adequately trained. In **chapter 11**, we outlined the current state of the medical curriculum, which is still mostly memorization based. We recommended changes as part of a staged approach, which can educate medical students on AI along their journeys. With appropriate training, future physicians will be prepared to work with the tools and techniques we implemented throughout this thesis.

*To end this thesis on a personal note, I want to finish it where it started by remembering hearing about the changing sepsis definition in 2016. I felt confused that the sepsis syndrome I had seen and learned about suddenly did not exist anymore. If anything, the last four years of research have puzzled me even more. The high complexity of the interaction between microbes and the host makes it so that researchers still struggle to define sepsis, even 5000 years after its symptoms have supposedly been first described. To accurately define and treat it, we would need a near-perfect understanding of the pathophysiology, which is still far out of sight. This thesis supports the bold prediction that data and machine learning hold the key to further unraveling the secrets of sepsis.*

## REFERENCES


1. Dellinger, R. P. The Future of Sepsis Performance Improvement. *Critical Care Medicine* vol. 43 1787–1789 Preprint at https://doi.org/10.1097/CCM.0000000000001231 (2015).
2. Evans, L. *et al.* Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* (2021) doi:10.1007/s00134-021-06506-y.
3. Damiani, E. *et al.* Effect of Performance Improvement Programs on Compliance with Sepsis Bundles and Mortality: A Systematic Review and Meta-Analysis of Observational Studies. (2015) doi:10.1371/journal.pone.0125827.
4. Coburn, B., Morris, A. M., Tomlinson, G. & Detsky, A. S. Does this adult patient with suspected bacteremia require blood cultures? *JAMA - Journal of the American Medical Association* **308**, 502–511 (2012).
5. Goto, M. & Al-Hasan, M. N. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clinical Microbiology and Infection* vol. 19 501–509 Preprint at https://doi.org/10.1111/1469-0691.12195 (2013).
6. Klucher, J. M., Davis, K., Lakkad, M., Painter, J. T. & Dare, R. K. Risk factors and clinical outcomes associated with blood culture contamination. *Infect Control Hosp Epidemiol* 1–7 (2021) doi:10.1017/ice.2021.111.
7. Marshall, J. C. Why have clinical trials in sepsis failed? *Trends Mol. Med.* **20**, 195–203 (2014).
8. Vincent, J.-L. Current sepsis therapeutics. *EBioMedicine* **0**, 104318 (2022).
9. Marshall, J. C. *et al.* A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases* vol. 20 e192–e197 Preprint at https://doi.org/10.1016/S1473-3099(20)30483-7 (2020).
10. Komorowski, M., Green, A., Tatham, K. C., Seymour, C. & Antcliffe, D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine* **0**, 104394 (2022).
11. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med* **34**, 1589–1596 (2006).
12. Densen, P. Challenges and Opportunities Facing Medical Education. *Trans Am Clin Climatol Assoc* **122**, 48 (2011).

# CHAPTER 14

## NEDERLANDSE SAMENVATTING

## SAMENVATTING

Dit laatste hoofdstuk geeft een samenvatting van de bevindingen in dit proefschrift. Mijn onderzoek heeft tot doel gehad om de basisprincipes van sepsis zorg te onderzoeken, inclusief de diagnostische work-up en behandeling, en manieren te vinden waarop data en machine learning deze kunnen verbeteren.

## SEPSIS PRESTATIEVERBETERPROGRAMMA'S

Om manieren te vinden waarop data en machine learning kunnen bijdragen aan de sepsis zorg, moeten we eerst stilstaan bij hoe die zorg er nu uit ziet. De internationale Surviving Sepsis Campaign (SSC) maakt al sinds haar oprichting in het begin van de eenentwintigste eeuw richtlijnen voor de behandeling van sepsis[1]. Om te zorgen dat deze adviezen ook geïmplementeerd worden, heeft de SSC de aanbevelingen gebundeld in overzichtelijke groepen van vergelijkbare interventies die voor een bepaalde tijd uitgevoerd moeten worden. Toch worden deze richtlijnen lang niet altijd gevolgd. Als reactie hierop hebben ziekenhuizen zogenaamde sepsis prestatieverbeterprogramma's opgezet, om de naleving van de richtlijnen te bevorderen. De laatste update van de SSC-richtlijn beveelt aan dat ziekenhuizen en zorgstelsels dergelijke sepsis prestatieverbeterprogramma's gebruiken[2]. **Hoofdstuk 2** bespreekt de beschikbare literatuur over het gebruik en de voordelen van prestatieverbeteringsprogramma's voor sepsis. We bieden ook praktische inzichten voor de klinische implementatie hiervan. In de literatuur worden programma's voor prestatieverbetering bij sepsis consistent geassocieerd met verbeterde naleving van de richtlijnen en met lagere sterftecijfers[3]. Idealiter integreren deze programma's screeningtools, veranderingen in sepsiszorgtrajecten en educatieve initiatieven om het bewustzijn over sepsiszorg te vergroten. Het betrekken van grote multidisciplinaire groepen belanghebbenden, waaronder patiënten, is essentieel om deze programma's met succes uit te voeren.

Gebruikmakend van wat we geleerd hebben over prestatieverbeterprogramma's voor sepsis, implementeerden we een dergelijke interventie in ons ziekenhuis in **hoofdstuk 3**. We voerden een voor-na-interventiestudie uit op de spoedeisende hulp van het Amsterdam UMC. De interventie bestond uit een screeningsinstrument, educatieve bijeenkomsten, audits en feedback, en een multidisciplinair sepsis team. De postinterventiefase ging gepaard met verbeteringen in de meeste procesgerelateerde uitkomsten, zoals een kortere tijd tot antibiotica (66 vs. 143 minuten; $p<0.001$), meer lactaatmetingen (72.9% vs. 46.2%; $p<0.001$), en meer voltooide Modified Early Warning Scores (MEWS; 85.0% vs. 62.9%; $p<0.001$) in vergelijking met de pre-interventiefase. Er waren echter geen verschillen in patiëntgerelateerde uitkomsten tussen de pre-interventie- en postinterventiefasen, met uitzondering van een verbeterd percentage van directe versus uitgestelde intensive care-opnames (100% direct vs. 64.3% direct; $p=0.012$). We concluderen dat het programma samenwerking en tijdige besluitvorming stimuleerde en de naleving van het protocol verbeterde, terwijl artsen nog steeds konden werken volgens hun klinische oordeel. De resultaten kunnen urgentie creëren voor een grotere RCT om de waarde van sepsis-prestatieverbeterprogramma's nog beter in kaart te brengen.

## DE DIAGNOSTISCHE WORK-UP VAN SEPSIS

Aangezien sepsis uit vele bronnen en ziekteverwekkers kan ontstaan, is de diagnostische work-up zeer divers en uitdagend om in een richtlijn vast te leggen. Kunstmatig intelligente (AI) tools kunnen meer op maat gemaakte aanbevelingen doen door grote hoeveelheden gegevens te doorzoeken en subtiele patronen te vinden die mensen mogelijk over het hoofd zien. In **hoofdstuk 4** hebben we beschikbare AI-beslissingsondersteunende tools voor sepsis in kaart gebracht om hun potentieel en valkuilen te verkennen. Door de kwaliteit van de opgenomen onderzoeken te beoordelen met behulp van de PROBAST-tool, realiseerden we ons dat sommige gevallen waarin AI voor sepsis werd gebruikt, aanzienlijke uitdagingen met zich meebrachten. AI-modellen die het begin van sepsis voorspelden, gebruikten vaak voorspellende variabelen zoals bloeddruk. Deze voorspeller maakt ook deel uit van de sepsis-definitie, die het probeert te voorspellen. Vanwege deze integratiebias worden de prestaties van deze tools overschat.

Een cruciaal onderdeel van het diagnostisch onderzoek waarvoor weinig AI-modellen bestaan, is het voorspellen van de resultaten van (bloed)kweken. Artsen vragen veel bloedkweken aan uit angst om bloedbaaninfecties te missen, die hoge morbiditeits- en mortaliteitscijfers hebben[4,5]. Het aanvragen van bloedkweken zonder de juiste indicatie leidt echter tot lage opbrengsten, terwijl de patiënt het risico loopt op schade door vals-positieve resultaten. Gecontamineerde bloedkweken (vals-positieve kweken) worden geassocieerd met aanvullende microbiologische testen, onnodig gebruik van antibiotica, langdurig ziekenhuisverblijf en zelfs ziekenhuissterfte[6]. **Hoofdstuk 5** laat zien hoe we een machine learning-algoritme hebben ontwikkeld om bloedkweekresultaten op de spoedeisende hulp te voorspellen, gevolgd door multicenter validatie en prospectieve evaluatie van de tool. De area under the curve (AUC) van de modelvoorspellingen of een bloedkweek positief zal worden was 0.81 (95%-BI = 0.78–0.83) in Amsterdam UMC locatie VU medisch centrum (VUmc) en tussen 0.75-0.80 in drie externe cohorten in Nederland en de Verenigde Staten. We hebben het algoritme verder geïmplementeerd in het elektronische patiëntendossier van VUmc voor real-time evaluatie, waarin het een AUC van 0.76 behield. Door de tool te gebruiken om bloedkweektesten bij patiënten met een laag risico achterwege te laten, kunnen we het aantal bloedkweken op de spoedeisende hulp mogelijk met 30% verminderen. Er moeten echter eerst verschillende uitdagingen worden aangepakt om de toepassing in de praktijk te laten slagen.

De belemmerende en bevorderende factoren die zorgprofessional ervaren bij klinische AI-implementatie zijn weinig beschreven. **Hoofdstuk 6** onderzoekt die barrières in een mixed-methods-onderzoek met interviews, focusgroepen en een landelijke vragenlijst. De belangrijke thema's die uit de discussies naar voren kwamen, zoals gevoel van noodzaak tot verandering, toegang tot kennis en informatie en bewijskracht, werden vervolgens gekoppeld aan geschikte implementatiestrategieën volgens de Expert Recommendations for Implementing Change (ERIC). We concluderen dat het huidige gevoel van noodzaak tot verandering om AI te implementeren moet worden aangewakkerd om duurzame implementatie mogelijk te maken. Dit kan worden bereikt met behulp van educatieve bijeenkomsten en betrokken lokale leiders. Professionals in de gezondheidszorg willen ook proeffasen om AI-aanbevelingen te vergelijken met hun eigen oordeel om het vertrouwen in de tools te vergroten. Ten slotte is het voor hen van cruciaal

belang dat AI-algoritmen zijn aangepast op de lokale context en passen in bestaande werkwijze. Met al deze aspecten in het achterhoofd, hebben we ons voorspelmodel voor bloedkweken verder geoptimaliseerd en een RCT opgezet om de potentiële voordelen ervan voor patiënten te onderzoeken. In de algemene discussie werd de studie verder toegelicht.

## HET OPTIMALISEREN VAN DE BEHANDELING VAN SEPSIS

Ondanks meer dan honderd RCT's naar immunomodulerende geneesmiddelen, bestaat de huidige behandeling van sepsis nog altijd alleen uit brede en niet-specifieke interventies zoals resuscitatie en infectiebestrijding[7,8]. De significante heterogeniteit van de sepsispopulatie is een primaire reden voor het ontbreken van positieve onderzoeksresultaten voor nieuwe behandelingen. Veel studies hebben nieuwe therapieën onderzocht bij patiënten met verschillende kenmerken en soorten infecties. Onlangs heeft het onderzoeksveld haar focus verlegd naar het vinden van subgroepen van patiënten met vergelijkbare kenmerken die mogelijk beter op behandelingen reageren. De pandemie van het coronavirus (COVID-19) versterkte onverwachts de waarde van deze aanpak. In **hoofdstuk 7** beschrijven we dat nieuwe en effectieve therapieën voor deze specifieke subgroep van virale sepsis binnen enkele maanden werden gevonden. Het hoofdstuk beschrijft verschillende aspecten van COVID-19-onderzoek waarvan sepsisonderzoekers kunnen leren, zoals het gebruik van een breder scala aan uitkomstmaten, wereldwijde samenwerking met gestandaardiseerde gegevensverzameling en het bestuderen van een meer homogene ziekte[9].

Aangezien COVID-19-onderzoek slechts één enkele ziekteverwekker beschouwt, is het wezenlijk anders dan het onderzoeken van de brede verzameling symptomen en uitkomsten onder de overkoepelende term sepsis[8,10]. Maar zelfs in COVID-19 is er een aanzienlijke heterogeniteit in hoe patiënten reageren op een infectie met SARS-CoV-2. In **hoofdstuk 8** hebben we clustertechnieken gebruikt om meer homogene subgroepen binnen de COVID-19-populatie te vinden. We repliceerden de methoden van een Franse studie om te zien of we vergelijkbare subfenotypes van COVID-19 konden vinden. Onze clusteranalyses op een multicenter Nederlands cohort vonden inderdaad vergelijkbare verdelingen van kenmerken. Subfenotype 1 bestond voornamelijk uit relatief jonge vrouwelijke patiënten (74.5%) met een hoge prevalentie van gastro-intestinale klachten maar weinig comorbiditeiten. Dit subfenotype had de gunstigste uitkomst. Subfenotype 2 bestond uit meer mannelijke patiënten (80.4%). Ze vertoonden minder symptomen maar hadden slechtere uitkomsten dan subfenotype 1. Subfenotype 3 bestond uit oudere, meestal mannelijke patiënten met verschillende comorbiditeiten. Dit subfenotype werd geassocieerd met de slechtste resultaten. Door het aantonen dat deze clusteranalyses relatief vergelijkbare subgroepen van COVID-19-patiënten opleveren als de Franse studie, ondersteunen de resultaten de robuustheid van de aanpak. Er waren echter ook duidelijke verschillen tussen de Franse en Nederlandse clusters, en we moeten voorzichtig blijven bij het gebruik van deze technieken om ziekteheterogeniteit te begrijpen.

Op basis van bovenstaande bevindingen wilden we dergelijke strategieën toepassen om groepen sepsispatiënten te ontdekken die mogelijk het meest baat hebben bij specifieke interventies. Een interessante interventie in dit opzicht is de vroege toediening van antibiotica voor sepsis, wat

mogelijk het dichtst in de buurt komt van een sepsis-specifieke therapie. Er is een diepgewortelde overtuiging dat elk uur uitstel van de toediening van antibiotica de overlevingskansen zal verkleinen, zoals werd beschreven in een veel geciteerd artikel uit 2006[11]. **Hoofdstuk 9** besprak de literatuur over de voordelen van vroege antibiotica voor sepsis. Het bewijs dat deze interventie ondersteunt is alleen gebaseerd op retrospectieve gegevens, terwijl meta-analyses en de enige gerandomiseerde klinische studie over dit onderwerp geen voordelen hebben laten zien. Toch vragen de huidige richtlijnen artsen vaak om patiënten met mogelijke sepsis zo snel mogelijk te behandelen met breedspectrumantibiotica. In de review stelden wij voor dat de echte uitdaging is om gevallen te identificeren waarin we de antibioticabehandeling veilig kunnen uitstellen om aanvullende gegevens te verzamelen om de waarschijnlijkheid van een bacteriële infectie te vergroten of te verkleinen. Hoewel deze algemene aanbeveling de behandeling van sepsis waarschijnlijk zal verbeteren, konden we nog steeds geen gemakkelijk identificeerbare subgroepen van patiënten vinden die, op groepsniveau, baat zouden hebben bij vroege antibioticabehandeling.

In **hoofdstuk 10** hebben we clusteranalyses gebruikt om te proberen die subgroepen te identificeren in de onderzoekspopulatie van de enige RCT over dit onderwerp, de pre-hospitale antibiotica tegen sepsis (PHANTASi) trial. Verrassend genoeg toonden de clusterpatronen consequent aan dat leeftijd veruit de belangrijkste factor was voor clustervorming. Toen we vervolgens de sterftecijfers over clusters en interventiegroepen samenvatten, was er een interactie tussen de voordelen van vroege antibioticabehandeling en jongere leeftijd. Verdere logistische regressiemodellering om deze relatie te kwantificeren, gecorrigeerd voor confounders, bevestigde een significante interactie tussen jongere leeftijd en de voordelen van vroege antibiotica. Na correctie voor deze interactie konden we een significant voordeel vinden van vroege antibioticabehandeling (odds-ratio 0.07; 95%-BI 0.01-0.79; p = 0.03) in de volledige onderzoekspopulatie. Met behulp van deze data-gedreven inzichten kunnen we werken aan meer gepersonaliseerde aanbevelingen voor sepsisbehandelingen.

## EPILOOG



Het zorglandschap evolueert en data en machine learning ter ondersteuning van medische besluitvorming zullen steeds belangrijker worden. In 2020 werd geschat dat de totale hoeveelheid medische informatie elke 73 dagen verdubbelde[12]. Nieuwe technologieën zoals AI kunnen het effectieve gebruik van al deze informatie mogelijk maken. Dit vereist bekwame interacties tussen mens en machine, waarvoor professionals in de gezondheidszorg voldoende moeten worden opgeleid. In **hoofdstuk 11** schetsten we de huidige stand van zaken in het medische curriculum, dat nog grotendeels gebaseerd is op het onthouden van informatie. We hebben wijzigingen aanbevolen als onderdeel van een gefaseerde aanpak, waarmee medische studenten tijdens hun onderwijs kennis kunnen opdoen over AI. Met de juiste training zullen toekomstige artsen voorbereid zijn om te werken met de tools en technieken die we in dit proefschrift hebben geïmplementeerd.

*Om dit proefschrift met een persoonlijke noot af te sluiten, wil ik het afmaken waar het mee begon. Ik herinner me weer dat ik hoorde over de veranderende sepsisdefinitie in 2016. Ik voelde me verward dat het sepsissyndroom dat ik had gezien en waarover ik had geleerd, plotseling niet*

*meer bestond. De afgelopen vier jaar onderzoek hebben me wellicht zelfs nog meer verward. De complexiteit van de interactie tussen microben en de mens zorgt ervoor dat onderzoekers nog steeds moeite hebben om sepsis te definiëren, zelfs 5000 jaar nadat de symptomen voor het eerst zijn beschreven. Om sepsis nauwkeurig te definiëren en te behandelen, zouden we een bijna perfect begrip van de pathofysiologie nodig hebben, wat nog ver uit het zicht is. Dit proefschrift ondersteunt de gedurfde voorspelling dat data en machine learning de sleutel vormen om de geheimen van sepsis verder te ontrafelen.*

## REFERENTIES


1. Dellinger, R. P. The Future of Sepsis Performance Improvement. *Critical Care Medicine* vol. 43 1787–1789 Preprint at https://doi.org/10.1097/CCM.0000000000001231 (2015).
2. Evans, L. *et al.* Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* (2021) doi:10.1007/s00134-021-06506-y.
3. Damiani, E. *et al.* Effect of Performance Improvement Programs on Compliance with Sepsis Bundles and Mortality: A Systematic Review and Meta-Analysis of Observational Studies. (2015) doi:10.1371/journal.pone.0125827.
4. Coburn, B., Morris, A. M., Tomlinson, G. & Detsky, A. S. Does this adult patient with suspected bacteremia require blood cultures? *JAMA - Journal of the American Medical Association* **308**, 502–511 (2012).
5. Goto, M. & Al-Hasan, M. N. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clinical Microbiology and Infection* vol. 19 501–509 Preprint at https://doi.org/10.1111/1469-0691.12195 (2013).
6. Klucher, J. M., Davis, K., Lakkad, M., Painter, J. T. & Dare, R. K. Risk factors and clinical outcomes associated with blood culture contamination. *Infect Control Hosp Epidemiol* 1–7 (2021) doi:10.1017/ice.2021.111.
7. Marshall, J. C. Why have clinical trials in sepsis failed? *Trends Mol. Med.* **20**, 195–203 (2014).
8. Vincent, J.-L. Current sepsis therapeutics. *EBioMedicine* **0**, 104318 (2022).
9. Marshall, J. C. *et al.* A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases* vol. 20 e192–e197 Preprint at https://doi.org/10.1016/S1473-3099(20)30483-7 (2020).
10. Komorowski, M., Green, A., Tatham, K. C., Seymour, C. & Antcliffe, D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine* **0**, 104394 (2022).
11. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med* **34**, 1589–1596 (2006).
12. Densen, P. Challenges and Opportunities Facing Medical Education. *Trans Am Clin Climatol Assoc* **122**, 48 (2011).

# CHAPTER 15

## ADDENDUM

# LIST OF CONTRIBUTING AUTHORS

*Center for Experimental and Molecular Medicine (CEMM), Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands*
Richarda Vleghels; Brent Appelman; Joe M. Butler; Alex R. Schuurman; Hessel Peters-Sengers; Harjeet S. Virk; Tom van der Poll; W. Joost Wiersinga

*Department of Internal Medicine, Amsterdam UMC Location VU University Amsterdam, Amsterdam, The Netherlands*
Anneroos W. Boerman; Rishi S. Nannan Panday; Ketan Paranjape; Bo Schouten; Petra van Pijkeren; Maureen Thodé; Kayla Brugman; Nadia Alam; Tanca C. Minderhoud; Prabath W.B. Nanayakkara

*Department of Internal Medicine, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands*
Frits Holleman

*Emergency Department, Amsterdam UMC Location VU University, Amsterdam, The Netherlands*
Mahi Dzelili

*Emergency Department, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands*
Milan L. Ridderikhof

*Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden*
Niclas Skyttberg

*Department of Computer Science, Faculty of Science, VU University, Amsterdam, The Netherlands*
Frank C. Bennis

*Department of Medical Microbiology and Infection Prevention, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands*
Rogier P. Schade

*Department of Clinical Chemistry, Amsterdam UMC Location VU University Amsterdam, Amsterdam, The Netherlands*
Robert de Jonge

*Department of Strategy & Development, Amsterdam UMC, Amsterdam, The Netherlands*
Marlou van Beneden

*Roche Diagnostics Corporation, Indianapolis, United States*
Justin Kundert

*Lee Kong Cian School of Medicine, Nanyang Technological University, Singapore, Singapore*
Josip Car

*Department of EVA Service Center, Amsterdam UMC, Amsterdam, The Netherlands*
Mei Lie

# LIST OF PUBLICATIONS

## Publications included in this thesis

**Schinkel M,** Nanayakkara PWB, Wiersinga WJ (2022) Sepsis Performance Improvement Programs: From Evidence Toward Clinical Implementation. Critical Care 2022 26:1 26:1–8

**Schinkel M**, Holleman F, Vleghels R, et al. The impact of a sepsis performance improvement program in the emergency department: a before-after intervention study [published online ahead of print, 2022 Nov 17]. *Infection*. 2022;1-10. doi:10.1007/s15010-022-01957-x

**Schinkel M,** Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB (2019) Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med 103488

**Schinkel M,** Boerman AW, Bennis FC, et al (2022) Diagnostic stewardship for blood cultures in the emergency department: A multicenter validation and prospective evaluation of a machine learning prediction tool. EBioMedicine 82:104176

Schouten, B., **Schinkel, M**., Boerman, A. W., Pijkeren, P. van, Thodé, M., Beneden, M. van, Panday, R. N., Jonge, R. de, Wiersinga, W. J., & Nanayakkara, P. W. B. (2022). Implementing artificial intelligence in clinical practice: a mixed-method study of barriers and facilitators. *Journal of Medical Artificial Intelligence*, *5*(0), 12–12. https://doi.org/10.21037/JMAI-22-71

**Schinkel M,** Nannan Panday RS, Joost Wiersinga W, Nanayakkara PWB (2020) Timeliness of antibiotics for patients with sepsis and septic shock. J Thorac Dis 12:S66–S71

**Schinkel M,** Virk HS, Nanayakkara PWB, van der Poll T, Wiersinga WJ (2021) What Sepsis Researchers Can Learn from COVID-19. Am J Respir Crit Care Med 203:125–127

**Schinkel M,** Appelman B, Butler J, Schuurman A, Joost Wiersinga W (2021) Association of clinical sub-phenotypes and clinical deterioration in COVID-19: further cluster analyses. Intensive Care Medicine 2021 1–3

**Schinkel M,** Paranjape K, Kundert J, Nannan Panday RS, Alam N, Nanayakkara PWB (2021) Towards Understanding the Effective Use of Antibiotics for Sepsis. Chest. https://doi.org/10.1016/j.chest.2021.04.038

Paranjape K, **Schinkel M,** Nannan Panday R, Car J, Nanayakkara P (2019) Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ 5:e16048

## Publications not included in this thesis

Nannan Panday RS, **Schinkel M,** Nutbeam T, Alam N, Nanayakkara PWB (2019) The effects of a single dose of paracetamol in a critical phase of sepsis: a sub-analysis of the PHANTASi trial. Eur J Intern Med 70:e7–e9

Boerman AW, **Schinkel M**, Meijerink L, et al. Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. *BMJ Open*. 2022;12(1):e053332. Published 2022 Jan 4. doi:10.1136/bmjopen-2021-053332

Paranjape K, **Schinkel M,** Nanayakkara P (2020) Short Keynote Paper: Mainstreaming Personalized Healthcare-Transforming Healthcare through new era of Artificial Intelligence. IEEE J Biomed Health Inform 1–1

**Schinkel M,** Panday RN, Nanayakkara P (2020) Blood Culture Results Before and After Antimicrobial Administration. Ann Intern Med 172:439

**Schinkel M,** Boerman AW, Nanayakkara PWB (2020) A Different Perspective on the Use of Sepsis Alert. Crit Care Med 48:E632–E633

Rothrock SG, Cassidy DD, Barneck M, et al (2020) Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. Ann Emerg Med. https://doi.org/10.1016/j.annemergmed.2020.04.042

Nannan Panday RS, **Schinkel M,** Nanayakkara PWB Septic patients with cancer: Do prehospital antibiotics improve survival? Do not forget the underlying status influence!

Paranjape K, **Schinkel M,** Hammer RD, Schouten B, Nannan Panday RS, Elbers PWG, Kramer MHH, Nanayakkara P (2021) The Value of Artificial Intelligence in Laboratory Medicine. Am J Clin Pathol 155:823–831

Ariës MJH, van den Bergh JP, Beudel M, et al (2021) [Clinical course of COVID-19 in the Netherlands: an overview of 2607 patients in hospital during the first wave]. Ned Tijdschr Geneeskd 165:

Veldhuis L, Ridderikhof ML, **Schinkel M,** et al (2021) Early warning scores to assess the probability of critical illness in patients with COVID-19. Emergency Medicine Journal 38:901–905

**Schinkel M,** Bergsma L, Veldhuis LI, Ridderikhof ML, Holleman F (2022) Comparing complaint-based triage scales and early warning scores for emergency department triage. Emergency Medicine Journal 0:1–6

Ottenhoff MC, Ramos LA, Potters W, et al (2021) Predicting mortality of individual patients with COVID-19: a multicentre Dutch cohort. BMJ Open. https://doi.org/10.1136/BMJOPEN-2020-047347

Schuurman AR, Reijnders TDY, Saris A, et al (2021) Integrated single-cell analysis unveils diverging immune features of COVID-19, influenza, and other community-acquired pneumonia. Elife. https://doi.org/10.7554/ELIFE.69661

Caniels TG, Bontjer I, van der Straten K, et al (2021) Emerging SARS-CoV-2 variants of concern evade humoral immune responses from infection and vaccination. Sci Adv. https://doi.org/10.1126/SCIADV.ABJ5365

Appelman B, van der Straten K, Lavell AHA, et al (2021) Time since SARS-CoV-2 infection and humoral immune response following BNT162b2 mRNA vaccination. EBioMedicine. https://doi.org/10.1016/J.EBIOM.2021.103589

**Schinkel M,** Paranjape K, Nanayakkara PWB (2022) Response. Chest 161:e195–e196

van Gils MJ, Lavell A, van der Straten K, et al (2022) Antibody responses against SARS-CoV-2 variants induced by four different SARS-CoV-2 vaccines in health care workers in the Netherlands: A prospective cohort study. PLoS Med 19:e1003991

**Schinkel M** (2022) Guest Editorial - A personal journey in Acute Medicine. Acute medicine 21:66–67

Grobben M, Juncker HG, van der Straten K, et al (2022) Decreased Passive Immunity to Respiratory Viruses through Human Milk during the COVID-19 Pandemic. Microbiol Spectr. https://doi.org/10.1128/SPECTRUM.00405-22

**Schinkel M,** Nannan Panday RS, Nanayakkara PWB (2022) What is the cause of increased mortality in normothermic patients with suspected infection? Eur J Emerg Med 29:316–317

Paranjape K, **Schinkel M,** Nanayakkara P, Elbers P, Kramer M (2021) Implementing Artificial Intelligence in Health Care: Data and Algorithm Challenges and Policy Considerations. Am J Biomed Sci Res 11:280–287

Schuurman AR, Reijnders TDY, van Engelen TSR, et al (2022) The host response in different aetiologies of community-acquired pneumonia. EBioMedicine. https://doi.org/10.1016/J.EBIOM.2022.104082

Lavell AHA, Sikkens JJ, Edridge AWD, et al (2022) Recent infection with HCoV-OC43 may be associated with protection against SARS-CoV-2 infection. iScience. https://doi.org/10.1016/J.ISCI.2022.105105

Sikkens JJ, Buis DTP, Peters EJG, et al (2021) Serologic Surveillance and Phylogenetic Analysis of SARS-CoV-2 Infection Among Hospital Health Care Workers. JAMA Netw Open 4:e2118554–e2118554

**Schinkel M**, Holleman F. In reply to: 'Use TRIPOD when validating clinical prediction models'. *Emerg Med J*. 2023;40(1):81-82. doi:10.1136/emermed-2022-212701

Turgman O, **Schinkel M**, Wiersinga WJ (2023) Host Response Biomarkers for Sepsis in the Emergency Room. Critical Care 2023 [accepted, to be published in March 2023]

# PORTFOLIO

PhD period: July 2019 – May 2023

PhD supervisors: prof. dr. W.J. Wiersinga, prof. dr. P.W.B. Nanayakkara, and dr. F. Holleman.

| | Year | ECTS |
|---|---|---|
| **1. PhD training** | | |
| **General courses** | | |
| AMC World of Science | 2019 | 0.2 |
| Practical Biostatistics | 2019 | 1 |
| e-BROK | 2020 | 1 |
| Infectious Diseases | 2019 | 1 |
| **Specific courses** | | |
| Data science professional certificate (Coursera, IBM) | 2021 | 2 |
| Master's in data science (University of Boulder Colorado) | 2021-2023 | 22 |
| **Seminars, workshops and master classes** | | |
| Yearly CEMMposium | 2019-2022 | 1 |
| **Oral presentations at (inter)national conferences** | | |
| World Academic Congress of Emergency Medicine 2021 (online edition; invited speaker) | 2021 | 0.2 |
| Florida State University Emergency Medicine Research Day 2022 (online edition; invited speaker) | 2022 | 0.2 |
| **Poster presentations at (inter)national conferences** | | |
| SAM on the Tyne; Newcastle, England | 2019 | 0.5 |
| XVIII$^{th}$ Congress of the European Shock Society; Chiania, Greece | 2019 | 0.5 |
| SAM Bristol; Bristol, England | | |
| European Congress of Clinical Microbiology & Infectious Diseases (ECCMID); | 2022 | 0.5 |
| Lissabon, Portugal | 2022 | 0.5 |
| Sepsis 2022; Barcelona, Spain | 2022 | 0.5 |
| **2. Teaching** | | |
| **Lecturing** | | |
| Master Personalized Medicine (3 lectures) | 2022 | 1 |
| Research minor, Bachelor of Medicine (5 lectures) | 2019-2022 | 2 |
| Pharmacology, Bachelor of Medicine (10 lectures and exams) | 2019 | 3 |
| **Tutoring, Mentoring** | | |
| Joost Wattel; master internship | 2020 | 1.5 |
| Lyfke Bergsma; master internship | 2021 | 1.5 |
| Kayla Brugman; master internship | 2022 | 1.5 |
| Arvind Nannan Panday; bachelor internship | 2019 | 0.5 |
| Vincent Boon; bachelor internship | 2022 | 0.5 |
| Xin Taphoorn; bachelor internship | 2023 | 0.5 |
| **3. Parameters of Esteem** | | |
| **Grants (with the team)** | | |
| NVIAG Wetenschapsbeurs (€5000) | 2020 | |
| APH QoC Innovation Grant (€21.500) | 2021 | |
| APH QoC Valorisation/Dissemination Grant (€2500) | 2021 | |

**PhD Portfolio.** (continued)

| | Year | ECTS |
|---|---|---|
| **Awards and Prizes** | | |
| People's poster prize at SAM on the Tyne | 2019 | |
| Best poster presentation at Sepsis 2022 (€500) | 2022 | |
| **Other** | | |
| Editorial Board Member of the Acute Medicine Journal | 2022 | |

## ACKNOWLEDGEMENTS - DANKWOORD

Het is onmogelijk om in een paar pagina's te uiten hoe dankbaar ik ben voor alle vriendschappen en steun die ik in de afgelopen jaren heb mogen ervaren. Als ik één les heb geleerd is het dat onderzoek en acute zorg allebei teamsporten zijn. Zonder de fantastische mensen om mij heen, was het nooit gelukt om aan dit mooie promotietraject te beginnen, laat staan het succesvol af te ronden met dit proefschrift.

Uiteraard kon dit proefschrift niet tot stand komen zonder mijn (co-)promotoren Prabath, Joost en Frits. Prabath, jou leerde ik als eerste kennen tijdens mijn wetenschappelijke stage. Serendipiteit. Elke dag weer ben ik onder de indruk van hoe jij met beperkte middelen, maar een eindeloos enthousiasme, ontzettend veel voor elkaar krijgt. Je bent arts, onderzoeker en artiest. Na vijf jaar weet ik nog altijd niet welk van de drie jij nu het meeste bent, dus concludeer ik dat je bovenal een intrigerende mentor bent voor mij. Vanaf dag één wist ik dat als jij iets beloofde, je die belofte ook altijd nakwam. Dat was eigenlijk alles wat ik nodig had, want het gaf me enorm veel vertrouwen dat het hoe dan ook goed zou komen. Je beloofde me om een sollicitatie bij Joost en Frits te regelen, zodat ik verder onderzoek naar sepsis kon doen. En zo geschiedde.

Joost, jouw positivisme is ongeëvenaard. Je ziet overal de mogelijkheden, bent bereid om met iedereen samen te werken en kunt altijd meteen tot de kern komen in elk van de vele uiteenlopende projecten die jij coördineert. Je stimuleerde mij altijd om de lat hoog te leggen, zonder ooit te veel van me te vragen. Dat is een unieke kwaliteit. Bovenal waardeer ik het feit dat jij op een bijna magische wijze iedereen te vriend en tevreden kunt houden. Je bent in vele opzichten hetzelfde als Prabath, maar door jullie verschillende kwaliteiten op onderzoeksgebied heb ik van jullie allebei veel nieuwe dingen kunnen leren.

Frits, jij was de perfecte copromotor voor mijn promotieteam. Je hebt vaak een duidelijke en onderbouwde mening en zult niet wegschrikken van een confrontatie als men tegenstrijdige ideeën heeft. Ik vind het ontzettend inspirerend dat je altijd opkomt voor je standpunt en niet een diplomatieke uitweg zult kiezen als jij overtuigd bent dat je idee goed is. Overigens ben je ook de eerste om het toe te geven als iemand anders een goed argument heeft om je van gedachte te veranderen. Die combinatie is ontzettend krachtig. Als ik je in twee woorden zou omschrijven dan zijn dat eerlijkheid en passie. Dank voor alle leerzame en interessante gesprekken tijdens onze werkoverleggen bij de Starbucks of tijdens de lunch. Ik hoop dat we die de komende jaren kunnen blijven voortzetten.

Naast mijn promotieteam wil ik ook alle leden van de commissie, Jacobien Hoogerwerf, Tom van der Poll, Alexander Vlaar, Hester Lingsma, Jan Prins en Paul Elbers, van harte bedanken voor de tijd die jullie genomen hebben om mijn proefschrift te lezen en beoordelen.

Roos en Rishi. Van alle collega's heb ik met jullie toch de meest speciale band gekregen. Daarom ben ik ontzettend blij dat jullie bij de verdediging van dit proefschrift mijn paranimfen zijn. Rishi, jij hebt mij "gescout" tijdens de examens van farmacologie. Als ik jou nooit ontmoet had, dan was ik niet gekomen waar ik nu ben. Ontzettend bedankt voor al je steun. Je hebt me laten zien hoe ik efficiënt kan werken en gaf me alle mogelijkheden om te starten als onderzoeker. Roos, jij bent net na mij gestart met je promotieonderzoek. Door de jaren hebben we onze krachten en

verschillende kwaliteiten gebundeld, waardoor we een heel sterk team werden en ongelofelijk mooie projecten hebben opgezet. Op de momenten waar ik het minste raad wist met hoe we verder moesten, kwam jij met de oplossingen.

Richarda, wij hebben ons samen ruim drie jaar ingezet voor het sepsis team. Zonder jou als partner bij dit project was het mij nooit gelukt om dit tot een succes te maken. Je was altijd enthousiast en stond overal open voor. Ik ben je ontzettend dankbaar voor alles wat je hebt gedaan en hoop je snel weer te zien op de SEH.

Ketan, you might be the brightest person I have ever met. When I still did my research internships, I had the amazing opportunity to meet with you and partner to work together on getting our Ph.D. I felt out of my depth working with a world-leading tech and digital health expert. However, you were always kind and humble and inspired me along the way. I am grateful for all the time you have invested in me and my career. Because of you, I started to pursue a master's degree in Data Science. Now that you have obtained the Ph.D., it is my turn to defend our work.

Gezien ik aan beide kanten van de Amstel heb gewerkt, heb ik het geluk gehad om van twee fantastische groepen collega's te mogen genieten. In het AMC werd ik aangespoord om groots te denken op het gebied van onderzoek, en dat vooral in teamverband te doen. Alex, Bas, Bob, Brent, Christine, Emma, Erik, Evelien, Floor, Harjeet, Hessel, Ilse, Jason, Jelmer, Joe, Justin, Liza, Lonneke, Magda, Marleen, Niels, Nora, Nurul, Oren, Paola, Stijn, Tjitske, Tom, Valentine, Vanessa en Xanthe, dank voor al jullie gezelligheid en drive om mezelf nog net iets meer te pushen. In het VUmc werd ik geïnspireerd om klinisch te denken en het onderzoek zo in te steken dat de patiënt er uiteindelijk het meeste aan zou hebben. Ayesha, Babiche, Bo, Carlijn, David, Eva, Hanneke, Jara, Jonne, Kaoutar, Karlinde, Lars, Marjolein, Maureen, Rashudy, Sheena, Siham, Tanca, en Wilmar, dank voor alle leuke onderzoek besprekingen, koffietjes en teamwork.

Dank aan alle vrienden en familie die mij over de jaren gesteund hebben bij alles wat ik heb gedaan. Kay, Tim, Wilmer, en Willem, als ik met jullie iets afspreek is het altijd een feest. Aan al mijn geneeskunde studiegenootjes, Akki, Daitlin, Erik, Floor, Gijs, Jarik, Lars, Manon, Osoul, Ruben, Sterre en Thomas, door jullie positivisme en gedrevenheid heb ik altijd meer uit mezelf kunnen halen. Ondanks dat ik jullie de afgelopen vier jaar niet zo vaak heb gezien als ik zou willen, hoop ik dat we elkaar de komende jaren weer veel gaan tegenkomen, zowel op werk als daarnaast. Aan mijn ooms en tantes, Hans, Jill, Bert en Dolly, dank voor jullie steun en liefde. Hans, jij inspireerde me om onderzoek de doen waarmee ik mensen zou kunnen helpen. Ik zal de rondleiding in jouw universiteit in North Dakota nooit vergeten. Bert, jij hebt me laten zien dat geen uitdaging te groot of te gek is.

Een speciaal dankwoord voor jou, Wim, omdat mijn studie en carrière nooit zo waren gelopen zonder jou. Ik kan me nauwelijks voorstellen hoe druk je al was, maar toch heb je een week de tijd genomen om een middelbare scholier rond te leiden en alles van het ziekenhuis te laten zien. De hartoperatie die ik toen mocht meemaken zal ik nooit vergeten en mijn interesse voor de geneeskunde was voor altijd gewekt.

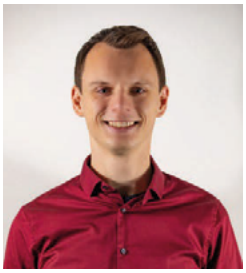
Alex, Angélique, Melissa en Ilona, wat heb ik een geluk dat jullie mijn (schoon)familie zijn. Vanaf het eerste moment dat ik bij jullie thuiskwam was ik altijd welkom en stonden jullie altijd voor mij klaar. Ik ben enorm dankbaar dat ik deel van jullie familie uit mag maken en jullie support en

gezelligheid heeft me de afgelopen jaar geholpen om altijd het beste uit mezelf te halen. Melissa, nog een extra bedankje voor het maken van de professionele foto voor in mijn proefschrift.

Mam, pap en Monique, bedankt voor jullie onvoorwaardelijke steun. Jullie staan altijd voor me klaar. Pap en mam, als ik even niet weet wat voor keuzes ik in mijn leven moet maken (moest ik bijvoorbeeld wel een promotieonderzoek gaan doen?), dan kom ik bij jullie voor advies. In plaats van jullie eigen mening, krijg ik dan altijd een spiegel voorgezet. Zo kwam ik erachter dat ik diep van binnen het antwoord meestal al wel wist. Jullie leerden het mij zelf te doen, waardoor ik me ironisch genoeg realiseer dat ik het eigenlijk nooit zelf doe. Er staat een team achter me. Ik hou van jullie en ben diep onder de indruk van hoe jullie als ouders een team zijn gebleven, zelfs nadat jullie uit elkaar zijn gegaan.

Lieve Louella. Jij bent de belangrijkste persoon in mijn leven. We leerden elkaar per toeval kennen op vakantie op Fuerteventura. Sindsdien realiseer ik me elke dag hoeveel geluk ik heb dat ik jou ben tegengekomen. Dat valt niet in woorden uit te drukken. Je hebt me gedurende dit promotietraject altijd gesteund. Je wist precies wanneer je me moest aanmoedigen om nieuwe uitdagingen aan te gaan en wanneer je even op de rem moest staan. Dat ik de afgelopen jaren zoveel plezier heb gehad in mijn onderzoek, studie en privéleven is bovenal aan jou te danken.

## ABOUT THE AUTHOR



Michiel Schinkel was born in Utrecht, the Netherlands. At fourteen years of age, Michiel had the amazing opportunity to spend a week in the "Onze Lieve Vrouwe Gasthuis (OLVG)" shadowing cardiothoracic surgeon Wim Stooker. From that moment, he knew he wanted to study medicine.

Throughout medical school, Michiel was interested in doing research. Through fortuitous meetings with Rishi, who conducted his pharmacotherapy examinations, Michiel got involved with research at the department of Acute Internal Medicine under the supervision of Prabath. Prabath introduced Michiel to Joost and Frits to interview for a Ph.D. trajectory on sepsis led by the three of them. On July 15th, 2019, this four-year journey started.

Through the amazing support of his family, friends, colleagues, and promotors, Michiel has had the fortunate chance to work on many interesting projects and explore how to shape his research career. As of January 1st, 2023, Michiel has published 36 scientific papers on various topics, such as sepsis, COVID-19, and artificial intelligence. He has also authored a chapter on artificial intelligence for sepsis in the "Sepsis Codex." His work has been cited 555 times, and Michiel won awards for best poster presentation at two international conferences. Because of his work, he was asked to be part of an expert panel to develop international recommendations on digital health competencies for undergraduate medical education in partnership with the World Health Organization.

Next to his research, Michiel's promotors encouraged him to keep studying and teaching. Since 2021, Michiel has been pursuing a master's degree in Data Science with the University of Boulder, Colorado, which he is expected to obtain in the first half of 2023. Subsequently, he has gotten involved in the Master Personalized Medicine at the Vrije Universiteit to give lectures and write assignments. Besides lecturing larger groups, Michiel has supervised three medical students during their bachelor's thesis and three during their research internships.

Michiel and his partner Louella recently moved from an apartment to their first house in Vianen, where they live happily. Michiel is an avid runner and enjoys racing any distance from 800 meters to the marathon. He also enjoys (slow-)cooking, bike riding, and playing board games with friends and family.