

# VU Research Portal

## Artificial Intelligence in Healthcare: Transitioning to Routine Clinical Care

Paranjape, Ketan

2022

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Paranjape, K. (2022). *Artificial Intelligence in Healthcare: Transitioning to Routine Clinical Care*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. Off Page.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### E-mail address:

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## **ARTIFICIAL INTELLIGENCE IN HEALTHCARE: TRANSITIONING TO ROUTINE CLINICAL CARE**

**Ketan Paranjape**

ISBN: 978-94-93278-13-4

Cover, layout and printing production: Off Page, Amsterdam

© K. Paranjape, United States of America, 2022

No part of this thesis may be reproduced, stored or transmitted in any form by any means  
without permission of the author.

VRIJE UNIVERSITEIT

**ARTIFICIAL INTELLIGENCE IN HEALTHCARE:  
TRANSITIONING TO ROUTINE CLINICAL CARE**

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan  
de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. J.J.G. Geurts,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de Faculteit der Geneeskunde  
op vrijdag 16 september 2022 om 13.45 uur  
in een bijeenkomst van de universiteit,  
De Boelelaan 1105

door

Ketan Paranjape  
geboren te Pune, India

**promotor:** prof.dr. K.D.P.W.B. Nanayakkara  
prof.dr. M.H.H. Kramer

**copromotoren:** dr. P.W.G. Elbers

**promotiecommissie:** prof.dr. R.M.C. Herings  
dr. J. Baron  
prof.dr. M. Hoogendoorn  
prof.dr. M. Bosch  
prof.dr. C. Boer  
prof.dr. R. de Jonge

For Anjali and Rishi



## TABLE OF CONTENTS

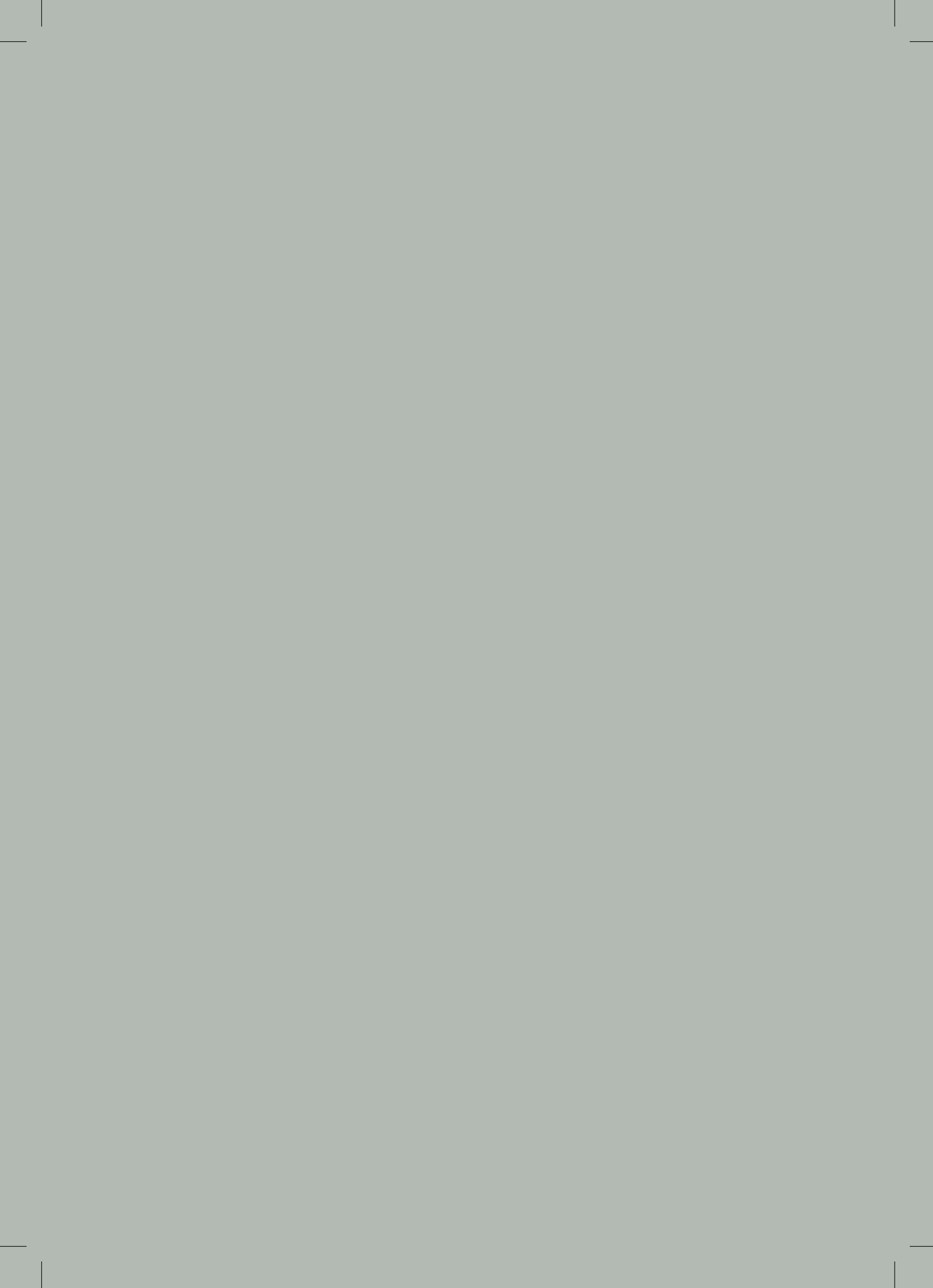
<b>Chapter 1</b>	General introduction and outline of this thesis	9
<b>Chapter 2</b>	Mainstreaming personalized healthcare – transforming healthcare through new era of artificial intelligence <i>IEEE J. Biomed. Health Info.</i> 2020 Jul;24(7):1860-1863	31
<b>Chapter 3</b>	Clinical applications of artificial intelligence in sepsis: a narrative review <i>Comput. Biol. Med.</i> 2019 Dec;115:103488	43
<b>Chapter 4</b>	Towards understanding the effective use of antibiotics for sepsis <i>Chest.</i> 2021 Oct;160(4):1211-21	65
<b>Chapter 5</b>	Development of a ‘meta-model’ to address missing data, predict patient-specific cancer survival and provide a foundation for clinical decision support <i>J. Am. Med Info Assoc.</i> 2021 Mar;28(3):605–15	85
<b>Chapter 6</b>	The value of artificial intelligence in laboratory medicine – current opinions and barriers to implementation <i>Am J Clin Pathol.</i> 2021 Jun;155(6):823-31	129
<b>Chapter 7</b>	Introducing artificial intelligence training in medical education <i>JMIR Med Educ.</i> 2019 Dec;5(2):e16048	155
<b>Chapter 8</b>	Implementing artificial intelligence in health care: data and algorithm challenges and policy considerations <i>Am J Biomed Sci Res.</i> 2021 Jan;11(4):280–7	177
<b>Chapter 9</b>	Summary and general discussion of main findings	193
<b>Appendix</b>	List of publications	221
	Author Affiliations	224
	Acknowledgements	225
	About the Author	227



# CHAPTER

1

GENERAL INTRODUCTION AND  
OUTLINE OF THIS THESIS



## INTRODUCTION ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) is intelligence displayed by machines, in contrast with human intelligence. AI attempts to approximate or translate human behavior and thought processes into algorithms or computer processes and embody them into computers or machines so that outputs of those mimic outputs which humans may produce on similar inputs. AI therefore makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks.

AI is becoming an increasingly advanced, sophisticated, and meaningful field, and its uses and implications are far reaching. AI's value proposition is the ability to process vast amounts of data, and then act on that data through computing abilities. In contrast to the human mind, computers are scalable and tireless. By processing data, the computer algorithms (algorithm is defined as set of rules for calculations or other problem-solving processes or operations) can 'learn' much faster than any human and develop what we understand as AI. AI can, for example, learn to identify potential diseases, treatment plans, and trends based on sifting through information and analyzing patient history, and provide recommendations to support, inform, and enable physician decision-making or make decisions on which human's act. Given these diverse use cases, AI is also referred to as Automated Intelligence or Augmented Intelligence.

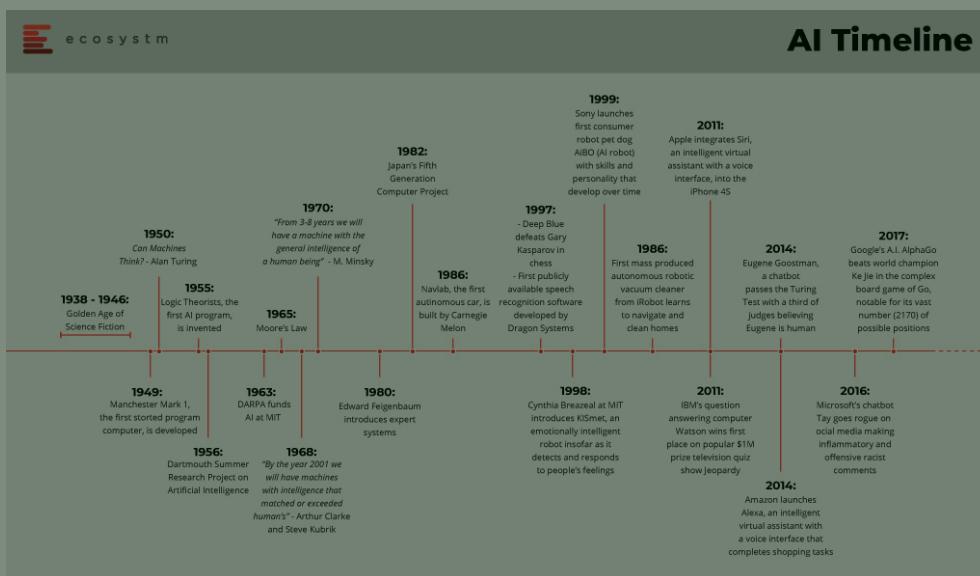
### Origins of AI

AI took roots through Sir Alan Turing's seminal paper *Computing Machinery and Intelligence* published in journal *Mind* in 1950 , where he proposed the question, "Can machines think?"(1).

The term Artificial Intelligence was coined by Dr. John McCarthy, who is known as the father of AI in 1956 after playing an influential role in defining the field devoted to the development of intelligent machines (2). His objective was to explore ways to make a machine that could reason like a human, was capable of abstract thought, problem-solving and self-improvement. He believed that "*every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.*"

### What is Driving the need for AI in Healthcare?

We need to understand the state of the healthcare today to truly appreciate what will drive AI. In many developed countries, mature but aged national healthcare services are being burdened with a growing aging population, accessibility to healthcare, changes in payment reforms, need for better diagnostics and treatments, worker shortage and rising costs of delivering care (3) (4) (5). Combined with a sudden surge in innovative technologies such as AI which can help with automating medical record to provide a more accurate diagnosis and tailored treatments, today's healthcare systems are ready for change (6). It is expected that the use of AI in healthcare global market is expected to grow from \$2.1 billion in 2018 to \$36.1 billion by 2025 (7).



**Figure 1.** History of Artificial Intelligence. Source - <https://blog.ecosystm360.com/5-ai-myths/>

Before we explore the wide world of AI applications in health and medicine, it is important to understand that AI is a constellation of numerous technologies.

## TYPES OF AI RELEVANT TO HEALTHCARE

AI is not a singular technology, rather a collection of them. Many of these technologies have demonstrated utility in healthcare and are described below.

### Machine Learning

Machine learning (ML) is a subfield of AI. Its goal is to facilitate computers to learn on their own. A ML algorithm enables the identification of patterns in data sets and predict results without having clear pre-programmed rules and models. This technique is at the core of many approaches to AI.

ML approaches in healthcare are used for predicting and recommending treatment options based on various patient characteristics and the treatment context (8). Majority of ML applications require a training dataset for which the outcome variable such as the onset of disease is known and this is called **supervised learning**.

Other applications of ML include early-stage drug discovery and finding alternative paths for therapy for multifactorial diseases using techniques such as next-generation sequencing. The ML technique involved here is called **unsupervised learning** which includes identifying patterns in data without providing any predictions.

The third kind of ML is **reinforcement learning** which aims to augment an individual's skills in the behavioral decision making through the use of experience of

the interaction with the world around them and create evaluative feedback. Contrary to supervised learning methods which relies on one-shot, comprehensive and supervised reward indicators, the reinforcement learning approach entails a progressive decision making that is simultaneously sampled, evaluative and comes with delayed feedback. This technology can predetermine ailments and treatments to help physicians and patients intervene at earlier stages.

## Neural Network

Neural networks (NN) are a class of algorithms that use a networked structure to learn from data. They are well suited to tackle problems such as prediction and pattern recognition (9). NN have been applied within medicine for clinical diagnosis (10), and more recently to image analysis (11) and interpretation, and drug development (12). This technique analyzes the problem in terms of inputs, outputs and weights of variables or 'features' that associate inputs with outputs. Also called Artificial Neural Networks (ANN), they are highly structured information processing units operating in parallel and attempt to mimic the huge computational ability of the human brain and nervous system.

## Deep Learning

Deep Learning (DL) can be regarded both as a sophisticated and mathematically complex evolution of machine learning algorithms. DL algorithms analyze data with logic similar to how a human would draw conclusions. DL includes both supervised and unsupervised learning. DL applications use a layered structure of algorithms called ANN as described above. The ANN is inspired by the natural neural network of the human brain. This leads to a far more capable process of learning compared to standard ML models.

One of the striking differences between DL and traditional ML is around the amount of data that is needed to train these systems. Traditional ML typically works with well-structured features and outputs that may encompass orders of magnitude fewer data points per case as compared to DL (e.g. tens to thousands of data points per case for traditional ML vs. sometimes millions for DL). Likewise, DL models often require a larger corpus to training cases to achieve adequate performance. One advantage of DL, is that it can often be applied directly to raw high dimensionality data (such as an image or block of text) without the need for extraction or curation of specific predefined features. In addition, while both traditional ML and DL can perform tasks such as classification or regression, some DL models, can actually generate new synthetic high dimensional outputs such as images, text or sounds. ML uses various types of automated algorithms that learn to model functions and predict future actions from data. DL uses neural networks that pass data through many processing layers to interpret data features and relationships.

One of the leading DL use cases in healthcare is around the recognition of potentially cancerous lesions in radiology images (13). Coupled with the ability to extract high

number of features from radiology images (Radiomics), DL is a very powerful tool for oncology image analysis (14).

### Natural Language Processing

Natural language processing (NLP) is the field of AI that is focused on the ability to understand and manipulate spoken or written data. NLP is useful for extracting information, unstructured to structured data conversion, document categorization and summarization.

In healthcare, NLP can be used for reducing administrative costs such as making billing efficient by extracting relevant information from unstructured physicians notes and appropriately assigning billing codes. NLP can also leverage information from physicians notes to alleviate delays and administrative errors to improve the prior authorization process. NLP can create medical value by helping healthcare providers with decision support at the point of need (e.g., predict post-surgery complications) and help compile and compare clinical guidance from various courses to identify the most appropriate care guidelines for care delivery.

Today NLP is used by voice activated assistants to enable patients' greater access to information and accurate transcription of patient – provider interaction resulting in better quality patient records. As an added benefit the provider gets more face time with the patient during the appointment and lowers stress (15).

### Cognitive Computing

Cognitive computing (CC) is the ability to simulate human thought processes via a computerized model. CC involves self-learning systems that use data mining, pattern recognition and NLP to mimic the way the human brain works. This technology is proving effective in applications such as clinical trial matching for lung, breast, colon and rectal cancer and analyzing electronic medical records for clinical efficiency and effectiveness at Mayo Clinic (16).

## AI APPLICATIONS IN HEALTH AND MEDICINE

Now with an understanding of AI technologies, let us consider how AI is being used in the patient and physicians' health journey. We live in a world of episodic care, where following an early or routine diagnosis, a physician develops a treatment plan based on a clinical decision support protocol. Then with the help of the health system, the physician gets paid for their services. Given the burdens we discussed earlier to all the healthcare stakeholders, we need to start focusing on wellness or preventative care and look at new options such as virtual assistants for triaging patients before they enter the hospital system. Finally, there are health conditions that go untreated resulting in an added burden to the health system.

Table 1 lists out AI applications in certain healthcare disciplines and Table 2 highlights AI use cases in select disease areas

**Table 1.** Summary of AI applications in healthcare

Key Focus Area	
<b>Patient Care</b>	
Assisted or automated diagnosis and participation	Chatbots can help patients self-diagnose or assist doctors in diagnosis (17) (18)
Prescription auditing	AI audit systems can help minimize prescription errors (19)
Pregnancy Management	Monitor mother and fetus to reduce mother's worries and enable early diagnosis (20)
Real-time prioritization and triage	Prescriptive analytics on patient data to enable accurate real-time case prioritization and triage (21)
Personalized medication and care	Find the best treatment plans according to patient data reducing cost and increasing effectiveness of care (22)
Patient data analytics	Analyze patient and/or 3rd party data to discover insights and suggest actions. AI allows the institution to analyze clinical data and generate deep insights into patient health. It provides an opportunity to reduce cost of care, use resources efficiently, and manage population health easily (23)
Wellness	AI can be used to learn from unlimited data from healthy individuals to provide detailed analytics on wellness management that is tailored to the individual (24) (25)
Surgical robots	Robot-assisted surgeries combines AI and collaborative robots. These robots perform repetitive tasks. AI can identify patterns within surgical procedures to improve best practices and to improve a surgical robots' control accuracy to sub-millimeter precision (26)
Preventative Care	Using data analysis to predict potential health outcomes, AI technology can help with diagnosis, recommend options for preventive care and improve accuracy.
<b>Clinical Decision Support, Medical Imaging and Diagnostics</b>	
Clinical Decision Support	Feeding well managed and curated health data to Machine Learning algorithms can enable efficient clinical decision support. This data can further be mined and cross-referenced with medical journals, data from other patients, and case studies to provide predictive analytics on the patient's health (27)
Early diagnosis	Analyze chronic conditions leveraging lab data and other medical data to enable early diagnosis (28)
Medical imaging analytics	Advanced medical imaging to analyze and transform images and model possible situations (29)

**Table 1.** (continued)

<b>Research and Development</b>	Drug discovery	Find new drugs based on previous data and medical intelligence (30)
	Clinical trial participation	Address concerns with patent cliff, analyzing real-world data and evidence, enabling outcome -driven approaches for treating conditions (31)
	Gene analysis and editing	Understand genes and their components. Predict the impact of gene edits (32)
	Device and drug comparative effectiveness	Applying AI to extract meaningful, actionable information from images and videos for experiment design (33)
<b>Healthcare Management</b>	Brand management and marketing	Create an optimal marketing strategy for the brand based on market perception and target segment (34)
	Pricing and risk	Determine the optimal price for treatment and other services according to competition and other market conditions (35)
	Market research	Prepare hospital competitive intelligence
	Operations	Process automation technologies such as intelligent automation and RPA help hospitals automate routine front office and back office operations such as reporting (36)
	Customer service chatbots	Customer service chatbots allow patients to ask questions regarding bill payment, appointments, or medication refills (18)
	Fraud detection	Patients may make false claims. Leveraging AI-powered fraud detection tools can help hospital managers to identify fraudsters (37)
	Cybersecurity	Ability to prevent breaches to protect patient data (38)

## CHALLENGES FOR AI IN HEALTH CARE

AI is becoming an increasingly advanced, sophisticated, and meaningful field, and its uses and implications are far-reaching. At the heart of AI's value proposition is the ability to process vast amounts of data and then act on that data through algorithms using techniques described above (47). One of the main challenges to implementing AI in health care are the role of algorithms and data.

Computer algorithms are at the core of AI and are part of applications in education, financial services, health care, navigation, and manufacturing. These algorithms are being used to make health care decisions such as prioritizing activities for staff members or triggering interventions for admitted patients—such as reported at John Hopkins Hospital(48). In these situations, how can one trust the algorithm to do the right thing,

**Table 2.** Examples of AI into various disease states

Disease Type	Clinical Management	AI Capability
Diabetic retinopathy (39)	Detection of early changes in fundi of patients with diabetes	Reading the retina and blood vessels to identify patients at risk of developing complicated diabetic retinal disease
Breast cancer (40)	Diagnosis of early breast cancer based on mammography	Reading mammographic pictures to detect early malignant transformation in breast cancer screening
Skin cancer (41)	Diagnosis of skin cancer by its clinical morphology	Identification of skin cancer by pictures and classification of types of skin neoplasia
Cerebrovascular disease (42)	Predicting outcome after a cerebrovascular accident	Predicting outcomes such as mortality of events such as stroke
Non-communicable chronic diseases (43)	Monitoring of diabetes and heart failure in primary care setting	Assisting patients monitoring of blood pressure and blood glucose at home and transmitting information to family medicine clinics
Heart Failures (44)	Predicting the clinical outcome of patients with heart failure	Predicting in-hospital mortality among patients with heart disease based on echocardiography
Sepsis (28)	Prediction models to diagnose sepsis	Assistance with treatment of sepsis, where the use of AI is associated with reducing mortality rates
Neurology (45)	Monitoring and management of neurodegenerative movement disorders	Detection and management of neurology conditions such as Parkinson's, Alzheimer's and Traumatic Brain Injury
Nephrology (46)	Patient management around prescriptions and transplants	Improve clinical care, hemodialysis prescriptions, and follow-up of transplant recipients

every time? For example, at Mount Sinai Hospital, when researchers applied deep learning to 700,000 patient records, without expert guidance the algorithm was able to identify patterns and predict the onset of diseases such as cancer of the liver. The algorithm also predicted the onset of schizophrenia but offered no clue as to how it did so. For a condition that is very difficult to predict even by the most experienced psychiatrist, the way the AI system came up with its decision is known as the "black box" problem(49). So how can we trust such a system?

Even more concerning are the consequences for the patient? If physicians tell their patients they are going to develop schizophrenia in the future and over time they do not develop it, what impacts does this have on the patients?

Decisions made by predictive algorithms can be obscure because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic

(the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations).

In considering the application of algorithms for individual patient treatment decisions, it should be noted that the physician will have to interpret the prediction model, as it is created by running the same algorithm on a mass scale. The physician will have to decide that the prediction score is reliable and accordingly propose a diagnosis.

AI relies both on the algorithm and patient data. To train a machine to identify specific conditions, hundreds of thousands of data elements are needed. For example, Google is using 46 billion data points to predict the clinical outcomes of hospital patients (50). The challenge here is that this need for data runs up against current models of patient privacy, consent, and control. For example, in the US, implementing AI in health care could violate HIPAA (Health Insurance Portability and Accountability Act) policies (51) and General Data Protection Regulation (GDPR) (52) (53).

Although HIPAA does not specifically address technologies such as AI, HIPAA was put in place to protect individuals' medical records and other personal health information. And there is a rapidly emerging move by consumers and patients to stop the "gold rush" mentality of industry to toss aside privacy and consent models (54). Patients are now understanding the importance of meaningful notice, consent, and control over their data, both HIPAA-covered data and data outside of HIPAA jurisdiction.

Within GDPR there are three provisions that enable patient privacy. First, is the European Union legislature explicitly addressing *algorithmic discrimination* by implementing technical and organizational measures to prevent discriminatory effects. Second is *data sanitization* that is the removal of special categories of datasets used in automated decision making. Third is *algorithmic transparency* that introduces the concept of "right to explanation" where meaningful information about the logic involved and the envisaged consequences when automated decision making or profiling takes place is shared with the data subjects (55).

When it comes to transitioning AI from research into routine clinical care, many challenges exist (56). First, while there are numerous studies that showcase the value of AI in healthcare with large number of patients and benchmarking against expert performance, most of them are retrospective studies with historically labelled data to train and test algorithms. There is a need for prospective studies on real-world data that differs from data used to train existing algorithms. Compounding this issue is the question about the quality of the data in the medical records. Data recorded in EMRs are affected by missing fields, medical codes, clinical protocols and health service capacity. There is a potential for these issues in the data set to impact the accuracy of the AI models (57,58).

Second, numerous AI studies are published on preprint servers and lack peer-review. There is a need to have more peer reviewed randomized controlled trials as an evidence gold standard.

Third, in order to understand the true potential of an AI algorithm, metrics such as area under the curve of a receiver operating characteristic curve (AUROC) (59), sensitivity and specificity are used today. In order to transition into clinical care, we

need metrics that quantify impact on patient care or how easily it can be applied to an existing clinical workflow (59).

Fourth, AI algorithms are susceptible to adversarial attacks or manipulation (60). Once deployed in clinical use, these models have to be protected and locked against such incidences.

We are therefore likely to encounter many ethical, medical, occupational and technological changes with AI in healthcare. Key healthcare stakeholders such as hospitals and regulatory agencies need a structure and framework to monitor issues with AI, react in a responsible and swift way and establish controls to limit the negative implications. This is one of the more powerful and consequential technologies to impact human societies, so it will require continuous attention and thoughtful policy for many years to come.

In order to take a deeper look into the implementation and mainstreaming of AI in healthcare, we decided to focus on two diseases – Sepsis and Cancer. Given the physiology, ability to diagnose and treat the disease, the impact on both the patient and healthcare organization in terms of cost of care and mortality, and finally the ability to collect and analyze plethora of data, we choose these two diseases.

## SEPSIS

Sepsis is a common but life-threatening disease that results in 49 million cases and 11 million annual deaths worldwide (61). In the US alone Sepsis is responsible for 270,000 annual deaths (3) and costs over \$27 billion in hospitalizations each year (62). Sepsis occurs when the body's response to an infection causes injury to its organs.

In the early stages of the disease, sepsis is relatively easy to treat with source control and broad spectrum antibiotics (63). 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 prognosis 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 (64). To improve patient outcomes, it is of the essence to improve time to diagnosis and accuracy of the prognosis for patients with sepsis.

By leveraging patient data, AI can help identify patients in the early stages of sepsis. Algorithms, dashboards displaying risk scores and automatic alerts can notify care teams about the progress of the patient and predict adverse events.

## CANCER

Cancer is the second leading cause of deaths in the world (65). 9.6 million people died due to various forms of cancer in 2018 (66). As the number of new cancer cases per year globally rises (expecting to reach 27.5 million by 2040) (65) and in conjunction as cancer

death rates decline in countries such as US (67), there is a need to learn more about this disease and advance cancer diagnostics methods and treatment options.

AI gives you the ability to sift through large volumes of medical data, extract relationships between various features and help identify characteristics in the data that may not be apparent to the human brain. Recent successes in radiology where algorithms processed images rapidly, augmented the radiologist's ability to make accurate decisions. In 2019, the Food and Drug Administration approved the first AI-based software to process images rapidly and assist radiologists in detecting breast cancer in screening mammograms.

By incorporating AI in cancer care, oncologist can improve the accuracy and speed of diagnosis, and augment clinical decision-making that could lead to better health outcomes. Algorithm driven care has the potential to play an important role in reducing healthcare costs along with improving accessibility and quality of care.

## AIMS AND OUTLINES OF THIS THESIS

The main aim of this thesis is to understand emerging technologies such as AI and how they are having an impact in healthcare. Our goal is to identify knowledge gaps and road blocks to the successful implementation and propose solutions to transition AI from research to routine clinical use.

In order to get a good understanding of AI there is a need to explain the technology and types of AI. We then need to understand what is driving the need for AI and what are the challenges in implementing AI in healthcare. Along with introducing the thesis, in **chapter 1** we addressed these topics and summarized results of a literature review of how AI is being used today in areas such as patient care, clinical decision support and imaging, healthcare management and R&D along with how AI is being used for clinical management of diseases such as sepsis and cancer.

Driven by data from modalities such as genomics, imaging and wearables, medicine is entering the digital age (68–74). As we gain a deeper understanding of the disease biology and how diseases affect an individual, we are developing targeted therapies to personalize treatments (75,76). There is a need for technologies such as AI to be able to support predictions for personalized treatments. In **chapter 2** we developed a consensus paper on how we can apply AI to the emerging field of personalized healthcare especially in sepsis and cancer given their disease burden.

Many studies have been published on a variety of clinical applications of AI for sepsis (77–86), but there is no systematic overview of the literature around applying AI methods for prognosis, predicting mortality and treatment of sepsis. In **chapter 3** we give an overview of this literature and identify knowledge gaps and prioritize areas for further research.

One such area is exploring the interaction between age and benefits of early antibiotics for sepsis (87–89) and if we can use AI techniques to identify any correlation. In **chapter 4** we applied AI to real world data from the only randomized clinical trial on this subject, the PreHospital Antibiotics Against Sepsis (PHANTASI) trial (90), to show that it is still plausible that subgroups of patients can benefit from this practice.

Similar to most real-world data, data derived from electronic medical records for patients typically exhibit wide inter-patient variability in terms of available data elements (91) (92) (93) (94) (27). This inter-patient variability leads to missing data and can present critical challenges in developing and implementing predictive AI models to underlie clinical decision support for patient-specific disease care. In **chapter 5** we proposed a novel ensemble approach to addressing missing data that we term the “meta-model” and apply the meta-model to patient specific disease prognosis. Using real-world data from advanced lung, colorectal and breast cancer, we developed a novel machine learning based strategy to underlie clinical decision support and predict survival in cancer patients, despite the missing data.

Having explored how AI is used in sepsis and cancer from a research standpoint, we sought to understand what are the challenges in transitioning AI from research to routine clinical care. Existing literature suggests that the main challenges include (56,95) - lack of accurate and sufficient data to develop and test AI models leading to increase bias in algorithms which is compounded by the lack of transparency in the models (black box effect) (42), privacy and control challenges over data (96), lack of knowledge about AI among the healthcare community (97) and the lack of policies and regulations around the use of AI from a patient care and liability standpoint (97). In the next three chapters we decided to first test these findings by conducting a survey of healthcare professionals. Second, we decided to understand the gaps in current medical education and propose a framework to introduce AI in medical education. Finally, we did a literature search of what polices and regulations are being developed and implemented to mainstream AI in routine clinical care.

In **chapter 6** we conducted a web-based survey on the use of AI in the laboratory medicine space. As digitization and automation proliferate in laboratory medicine, laboratorians will be faced with challenges associated with implementing and using technologies such as AI. There is going to be a need to get a good understanding of how to evaluate and implement AI. With this study we planned to evaluate the thoughts of laboratorians on the value of AI in the diagnostics space and identify anticipated challenges and solutions to introducing them.

Physicians go through extensive periods of training before they can eventually register as specialists. In spite of advances and changes in medicine over the past decade, traditional medical curricula has largely stayed the same (98). 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 (99). Introducing medical students and residents to new technologies such as telemedicine, mobile healthcare applications, robotics and AI is slowly developing (98) (99) (100). In **chapter 7** we conducted a literature study of AI training in medical education, clinical curriculum and continuous medical education. We discussed the findings and proposed a training framework to incorporate AI from standardized tests to medical school and specialty training.

Along with impressive advances and the realistic potential to transform healthcare in the near term there are difficult questions about liability and accountability, algorithmic bias and representative data, and the ability to accurately interpret and explain data that needs to be addressed to mainstream AI (101) (102,103). One key area to address is policy and regulation. In **Chapter 8** we addressed the challenges of implementing AI in routine health care practice by looking at the role of data and algorithms and the implications for policy, regulation and medical malpractice.

We aimed to understand emerging technologies such as AI and how they are having an impact in healthcare. We summarized and discussed our main findings in **Chapter 9**.

## REFERENCES

1. Alan M. Turing. Computing machinery and intelligence. *Mind*. 1950;
2. McCarthy J. What is artificial intelligence [Internet]. URL: <http://www-formal.stanford.edu/jmc/whatisai>. .... 2007. p. 1–15. Available from: <http://www-formal.stanford.edu/jmc/whatisai.pdf>
3. Vogenberg FR, Santilli J. Healthcare Trends for 2018. Am Heal drug benefits [Internet]. 2018 Feb [cited 2021 Feb 15];11(1):48–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29692880>
4. 2020 Global health care sector outlook | Deloitte [Internet]. [cited 2021 Feb 15]. Available from: <https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html>
5. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: Systematic review [Internet]. Vol. 22, *Journal of Medical Internet Research*. JMIR Publications Inc.; 2020 [cited 2021 Apr 22]. Available from: [/pmc/articles/PMC7059082/](https://pmc/articles/PMC7059082/)
6. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Futur Healthc J* [Internet]. 2019 Jun [cited 2021 Feb 15];6(2):94–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31363513>
7. Artificial Intelligence in Healthcare Spending to Hit \$36B [Internet]. [cited 2021 Apr 22]. Available from: <https://healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b>
8. Lee S-I, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun* [Internet]. 2018 Dec 3 [cited 2021 Mar 18];9(1):42. Available from: <http://www.nature.com/articles/s41467-017-02465-5>
9. Neural Networks in Healthcare [Internet]. [cited 2017 Jul 28]. Available from: <http://www.openclinical.org/neuralnetworksinhealthcare.html>
10. Patil S, Henry JW, Rubenfire M, Stein PD. Neural network in the clinical diagnosis of acute pulmonary embolism. *Chest* [Internet]. 1993 Dec [cited 2017 Jul 28];104(6):1685–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8252942>
11. Shi Z, He L, Suzuki K, Nakamura T, Itoh H. Survey on Neural Networks Used for Medical Image Processing. *Int J Comput Sci* [Internet]. 2009 Feb [cited 2017 Jul 28];3(1):86–100. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26740861>
12. Baskin II, Winkler D, Tetko I V. A renaissance of neural networks in drug discovery. *Expert Opin Drug Discov* [Internet]. 2016 Aug 2 [cited 2017 Jul 28];11(8):785–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27295548>
13. Fakoor R, Ladakh F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. undefined [Internet]. 2013 [cited 2021 Mar 18]; Available from: <https://www.semanticscholar.org/paper/Using-deep-learning-to-enhance-cancer-diagnosis-and-Fakoor-Ladakh/1e5ea08de7fc513c0958f9ab5b09266f60ca7c78>
14. Vial A, Stirling D, Field M, Ros M, Ritz C, Carolan M, et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Transl Cancer Res* [Internet]. 2018 [cited 2021 Mar 18];7(3). Available from: <https://tcr.amegroups.com/article/view/21823>
15. Craft JA. Artificial Intelligence and the Softer Side of Medicine. *Mo Med* [Internet]. 2018 [cited 2021 May 10];115(5):406–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30385982>
16. Artificial Intelligence Gets Real - Giving to Mayo Clinic [Internet]. [cited 2017 Aug 31]. Available from: <http://www.mayoclinic.org/giving-to-mayo-clinic/your-impact/features-stories/artificial-intelligence-gets-real>
17. Vaidyam AN, Wisniewski H, Halama JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can J Psychiatry* [Internet]. 2019 Mar 21 [cited 2019 Oct 13];64(7):070674371982897. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30897957>

18. Yaghoubzadeh R, Kramer M, Pitsch K, Kopp S. Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People. Proc 13th Int Conf Intell Virtual Agents [Internet]. 2013;8108:91. Available from: <http://pub.uni-bielefeld.de/publication/2623486>
19. Corny J, Rajkumar A, Martin O, Dode X, Lajonchère J-P, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Informatics Assoc [Internet]. 2020 Nov 1 [cited 2021 Feb 16];27(11):1688–94. Available from: <https://academic.oup.com/jamia/article/27/11/1688/5912186>
20. Iftikhar P, Kuypers M V, Khayyat A, Iftikhar A, DeGouvia De Sa M. Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice. Cureus [Internet]. 2020 Feb 28 [cited 2021 Feb 16];12(2):e7124. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32257670>
21. Wang M, Xia C, Huang L, al. et. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. Lancet Digit Heal. 2020;2:e506–15.
22. Freedman DH. Hunting for New Drugs with AI. Nature [Internet]. 2019 Dec 19 [cited 2021 Feb 16];576(7787):S49–53. Available from: <http://www.nature.com/articles/d41586-019-03846-0>
23. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. J Biomed Inform [Internet]. 2019 Dec 1 [cited 2021 Feb 16];100:103311. Available from: <https://www.sciencedirect.com/science/article/pii/S1532046419302308>
24. App abandonment is on the rise as consumers stick to the apps they know | TechCrunch [Internet]. [cited 2018 Jan 3]. Available from: <https://techcrunch.com/2016/11/18/app-abandonment-is-on-the-rise-as-consumers-stick-to-the-apps-they-know/>
25. Tran V-T, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. npj Digit Med [Internet]. 2019 Dec 14 [cited 2019 Sep 30];2(1):53. Available from: <http://www.nature.com/articles/s41746-019-0132-y>
26. Guo J, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. Heal equity [Internet]. 2018 [cited 2019 Oct 13];2(1):174–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30283865>
27. Baron JM, Kurant DE, Dighe AS. Machine Learning and Other Emerging Decision Support Tools. Clin Lab Med [Internet]. 2019 Jun [cited 2020 Mar 8];39(2):319–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31036284>
28. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med [Internet]. 2019 Oct 7 [cited 2019 Oct 16];103488. Available from: <https://www.sciencedirect.com/science/article/pii/S0010482519303567>
29. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. RadioGraphics [Internet]. 2017 Mar [cited 2019 Jul 28];37(2):505–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28212054>
30. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov [Internet]. 2019 [cited 2020 Feb 28];18(6):463–77. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30976107>
31. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci [Internet]. 2019 Aug 1 [cited 2021 Feb 16];40(8):577–91. Available from: <https://www.sciencedirect.com/science/article/pii/S0165614719301300>
32. Chen X, Ishwaran H. Pathway hunting by random survival forests. Bioinformatics [Internet]. 2013 Jan 1 [cited 2020 Mar 8];29(1):99–105. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23129299>
33. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open [Internet]. 2020 Oct 1 [cited 2021 Feb 16];3(3):326–31. Available from: <https://academic.oup.com/jamiaopen/article/3/3/326/5902619>

34. Uses For AI In Health Care Marketing [Internet]. [cited 2021 Feb 16]. Available from: <https://www.forbes.com/sites/forbesagencycouncil/2019/10/01/uses-for-ai-in-health-care-marketing/?sh=147261bd6f40>
35. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. *J Med Internet Res* [Internet]. 2020 [cited 2021 Feb 16];22(2):e16866. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32130134>
36. A look at operational and clinical applications of AI in healthcare - MedCity News [Internet]. [cited 2021 Feb 16]. Available from: <https://medcitynews.com/2019/07/operational-and-clinical-applications-of-ai-in-healthcare/>
37. Bughin J, Brussels J, Hazan E, Paris J, Ramaswamy S, Washington J, et al. ARTIFICIAL INTELLIGENCE THE NEXT DIGITAL FRONTIER? [Internet]. [cited 2018 Aug 15]. Available from: [www.mckinsey.com/mgi](http://www.mckinsey.com/mgi).
38. The Impact of AI on Cybersecurity | IEEE Computer Society [Internet]. [cited 2021 Feb 16]. Available from: <https://www.computer.org/publications/tech-news/trends/the-impact-of-ai-on-cybersecurity>
39. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* [Internet]. 2016 [cited 2021 Feb 16];316(22):2402–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27898976>
40. Rodriguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology* [Internet]. 2019 [cited 2021 Feb 16];290(2):305–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30457482>
41. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* [Internet]. 2017 [cited 2021 Feb 16];542(7639):115–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28117445>
42. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. *PLoS One* [Internet]. 2020 [cited 2021 Feb 16];15(4):e0231166. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32251471>
43. Ciccone MM, Aquilino A, Cortese F, Scicchitano P, Sassara M, Mola E, et al. Feasibility and effectiveness of a disease and care management model in the primary health care system for patients with heart failure and diabetes (Project Leonardo). *Vasc Health Risk Manag* [Internet]. 2010 May 6 [cited 2021 Feb 16];6:297–305. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20479952>
44. Kwon J-M, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography* [Internet]. 2019 [cited 2021 Feb 16];36(2):213–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30515886>
45. Belić M, Bobić V, Badža M, Šolaja N, Đurić-Jovičić M, Kostić VS. Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease-A review. *Clin Neurol Neurosurg* [Internet]. 2019 Sep [cited 2021 Feb 16];184:105442. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31351213>
46. Niel O, Bastard P. Artificial Intelligence in Nephrology: Core Concepts, Clinical Applications, and Perspectives. *Am J Kidney Dis* [Internet]. 2019 [cited 2021 Feb 16];74(6):803–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31451330>
47. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* [Internet]. 2018 Aug 1 [cited 2018 Nov 1];33(8):2358–61. Available from: <https://www.sciencedirect.com/science/article/pii/S0883540318302158>
48. The Johns Hopkins Hospital Launches Capacity Command Center to Enhance Hospital Operations - 10/26/2016 [Internet]. [cited 2018 Oct 14]. Available from: [https://www.hopkinsmedicine.org/news/media/releases/the\\_johns\\_hopkins\\_hospital\\_launches\\_capacity\\_command\\_center\\_to\\_enhance\\_hospital\\_operations](https://www.hopkinsmedicine.org/news/media/releases/the_johns_hopkins_hospital_launches_capacity_command_center_to_enhance_hospital_operations)

49. The Dark Secret at the Heart of AI - MIT Technology Review [Internet]. [cited 2017 Dec 23]. Available from: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>
50. Google is using 46 billion data points to predict a hospital patient's future — Quartz [Internet]. [cited 2018 Nov 1]. Available from: <https://qz.com/1189730/google-is-using-46-billion-data-points-to-predict-the-medical-outcomes-of-hospital-patients/>
51. Privacy | HHS.gov [Internet]. [cited 2017 Dec 24]. Available from: <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html>
52. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. *J Med Internet Res* [Internet]. 2017 [cited 2021 Mar 18];19(2):e47. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28235748>
53. Spencer A, Patel S. Applying the Data Protection Act 2018 and General Data Protection Regulation principles in healthcare settings. *Nurs Manag (Harrow)* [Internet]. 2019 Jan 28 [cited 2021 Mar 18];26(1):34–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31468806>
54. 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* [Internet]. 2018 [cited 2018 Oct 21];9:167. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29910822>
55. Forcier MB, Gallois H, Mullan S, Joly Y. Integrating artificial intelligence into health care through data access: Can the GDPR act as a beacon for policymakers? [Internet]. Vol. 6, *Journal of Law and the Biosciences*. Oxford University Press; 2019 [cited 2021 Apr 23]. p. 317–35. Available from: [/pmc/articles/PMC6813940/](https://doi.org/10.1093/jlb/lsz011)
56. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence [Internet]. Vol. 17, *BMC Medicine*. BioMed Central Ltd.; 2019 [cited 2021 Apr 26]. Available from: [/pmc/articles/PMC6821018/](https://doi.org/10.1186/s12881-019-0700-0)
57. Ta CN, Weng C. Detecting systemic data quality issues in electronic health records. In: *Studies in Health Technology and Informatics* [Internet]. IOS Press; 2019 [cited 2021 May 13]. p. 383–7. Available from: [/pmc/articles/PMC6857180/](https://doi.org/10.3233/SHTI1807180)
58. Stanfill MH, Marc DT. Health Information Management: Implications of Artificial Intelligence on Healthcare Data and Information Management [Internet]. Vol. 28, *Yearbook of medical informatics*. NLM (Medline); 2019 [cited 2021 May 13]. p. 56–64. Available from: [/pmc/articles/PMC6697524/](https://doi.org/10.1007/978-3-662-58870-0_6)
59. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. Brock G, editor. *PLoS One* [Internet]. 2015 Mar 4 [cited 2021 Apr 30];10(3):e0118432. Available from: [https://dx.plos.org/10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)
60. Finlayson SG, Chung HW, Kohane IS, Beam AL. Adversarial Attacks Against Medical Deep Learning Systems. 2018 Apr 14 [cited 2019 Feb 3]; Available from: [http://arxiv.org/abs/1804.05296](https://arxiv.org/abs/1804.05296)
61. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet (London, England)* [Internet]. 2020 Jan 18 [cited 2020 Feb 11];395(10219):200–11. Available from: [http://www.ncbi.nlm.nih.gov/pubmed/31954465](https://www.ncbi.nlm.nih.gov/pubmed/31954465)
62. Bloomberg: America Has a \$27 Billion Sepsis Crisis — Global Sepsis Alliance [Internet]. [cited 2021 Feb 15]. Available from: <https://www.global-sepsis-alliance.org/news/2017/7/17/bloomberg-america-has-a-27-billion-sepsis-crisis>
63. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign. *Crit Care Med*. 2013 Feb;41(2):580–637.
64. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019 May;321(20):2003.

65. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017 Sep 16;390(10100):1151–210.
66. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Heal* [Internet]. 2016 Sep [cited 2020 Mar 14];4(9):e609–16. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2214109X16301437>
67. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* [Internet]. 2019 Jan 8 [cited 2020 Mar 14];69(1):7–34. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21551>
68. Adamo JE, Bienvenu li RV, Dolz F, Liebman M, Nilsen W, Steele SJ. Translation of Digital Health Technologies to Advance Precision Medicine: Informing Regulatory Science. *Digit biomarkers* [Internet]. 2020 Feb 7 [cited 2021 Apr 26];4(1):1–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32399511>
69. Chen YW, Nishikawa I, Tamura S, Lu BL, Jiang H. Computational intelligence in biomedical science and engineering. Vol. 2012, *Computational Intelligence and Neuroscience*. 2012.
70. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Vol. 286, *Radiology*. Radiological Society of North America Inc.; 2018. p. 800–9.
71. Sacha GM, Varona P. Artificial intelligence in nanotechnology. Vol. 24, *Nanotechnology*. 2013.
72. Mainali G. Artificial intelligence in medical science: perspective from a medical student. Vol. 58, *Journal of the Nepal Medical Association*. Nepal Medical Association; 2020. p. 709–71.
73. Kritz F. Technology is transforming work for nurses and care for patients. Vol. 119, *American Journal of Nursing*. Lippincott Williams and Wilkins; 2019. p. 18–9.
74. Medicine in the digital age [Internet]. Vol. 25, *Nature Medicine*. Nature Publishing Group; 2019 [cited 2021 Apr 26]. Available from: <https://pubmed.ncbi.nlm.nih.gov/30617338/>
75. Krzyszczuk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized medicine for cancer treatment. *TECHNOLOGY* [Internet]. 2018 Sep [cited 2021 Apr 26];06(03n04):79–100. Available from: <https://pubmed.ncbi.nlm.nih.gov/30713991/>
76. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress [Internet]. Vol. 109, *Fertility and Sterility*. Elsevier Inc.; 2018 [cited 2021 Apr 26]. p. 952–63. Available from: [/pmc/articles/PMC6366451/](https://pmc/articles/PMC6366451/)
77. Sun Y, Kaur R, Gupta S, Paul R, Das R, Cho SJ, et al. Development and validation of high definition phenotype-based mortality prediction in critical care units. *JAMIA open* [Internet]. 2021 Jan 1 [cited 2021 Apr 26];4(1):ooab004. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33796821>
78. 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* [Internet]. 2018 Apr 1 [cited 2021 Apr 26];46(4):547–53. Available from: <https://pubmed.ncbi.nlm.nih.gov/29286945/>
79. Shashikumar SP, Josef CS, Sharma A, Nemati S. DeepAISE – An interpretable and recurrent neural survival model for early prediction of sepsis. *Artif Intell Med*. 2021 Mar 1;113.
80. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, 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 Jan 1;8(1).
81. Johnson AEW, Aboab J, Raffa JD, Pollard TJ, Deliberato RO, Celi LA, et al. A Comparative Analysis of Sepsis Identification Methods in an Electronic Database. *Crit Care Med*. 2018 Apr 1;46(4):494–9.
82. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Vol. 46, *Intensive Care Medicine*. Springer; 2020. p. 383–400.

83. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc.* 2018 Jul 1;15(7):846–53.
84. Wong A-KI, Cheung PC, Kamaleswaran R, Martin GS, Holder AL. Machine Learning Methods to Predict Acute Respiratory Failure and Acute Respiratory Distress Syndrome. *Front big data* [Internet]. 2020 Nov 23 [cited 2021 Apr 26];3:579774. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33693419>
85. Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. Vol. 33, *Journal of Clinical Monitoring and Computing*. Springer Netherlands; 2019. p. 887–93.
86. Ding M, Luo Y. Unsupervised phenotyping of sepsis using nonnegative matrix factorization of temporal trends from a multivariate panel of physiological measurements. *BMC Med Inform Decis Mak.* 2021 Apr;21(S5).
87. Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. *N Engl J Med.* 2017 Jun 8;376(23):2223–34.
88. Singer M. Antibiotics for sepsis: Does each hour really count, or is it incestuous amplification? [Internet]. Vol. 196, *American Journal of Respiratory and Critical Care Medicine*. American Thoracic Society; 2017 [cited 2021 Apr 26]. p. 800–2. Available from: [www.atsjournals.org](http://www.atsjournals.org).
89. 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 Oct 1 [cited 2021 Apr 26];196(7):856–63. Available from: [/pmc/articles/PMC5649973/](http://pmc/articles/PMC5649973/)
90. 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 1 [cited 2020 Aug 29];6(1):40–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29196046>
91. Weber GM, Adams WG, Bernstam E V, Bickel JP, Fox KP, Marsolo K, et al. Biases introduced by filtering electronic health records for patients with "complete data". *J Am Med Inform Assoc* [Internet]. 2017 Nov 1 [cited 2020 Nov 28];24(6):1134–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29016972>
92. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* [Internet]. 2013 Aug 1 [cited 2020 Mar 8];3(8):e002847. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23906948>
93. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. *Am J Clin Pathol* [Internet]. 2016 Jun [cited 2019 Aug 9];145(6):778–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27329638>
94. Luo Y, Szolovits P, Dighe AS, Baron JM. 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. *J Am Med Inform Assoc* [Internet]. 2018 [cited 2020 Nov 28];25(6):645–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29202205>
95. Park CW, Seo SW, Kang N, Ko BS, Choi BW, Park CM, et al. Artificial Intelligence in Health Care: Current Applications and Issues [Internet]. Vol. 35, *World Journal of Orthopedics*. Baishideng Publishing Group Co; 2020 [cited 2021 Apr 26]. p. 1–11. Available from: [/pmc/articles/PMC7606883/](http://pmc/articles/PMC7606883/)
96. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. *Front Public Heal* [Internet]. 2016 Feb 17 [cited 2018 Sep 15];4:7. Available from: <http://journal.frontiersin.org/Article/10.3389/fpubh.2016.00007/abstract>
97. Hazarika I. Artificial intelligence: Opportunities and implications for the health workforce. *Int Health* [Internet]. 2020 [cited 2021 Apr 26];12(4):241–5. Available from: [/pmc/articles/PMC7322190/](http://pmc/articles/PMC7322190/)
98. Pfeifer CM. A progressive three-phase innovation to medical education in the United States. *Med Educ Online.* 2018;00.

99. Nasca TJ, Philibert I, Brigham T, Flynn TC. The Next GME Accreditation System — Rationale and Benefits. *N Engl J Med.* 2012 Mar;366(11):1051–6.
100. Wartman SA, Donald Combs C. Medical education must move from the information age to the age of artificial intelligence. *Acad Med [Internet].* 2018 [cited 2021 Apr 26];93(8):1107–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/29095704/>
101. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: *Artificial Intelligence in Healthcare [Internet].* Elsevier; 2020 [cited 2021 Apr 26]. p. 295–336. Available from: [/pmc/articles/PMC7332220/](https://pmc/articles/PMC7332220/)
102. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: Accountability and safety. *Bull World Health Organ [Internet].* 2020 Apr 1 [cited 2021 Apr 26];98(4):251–6. Available from: [/pmc/articles/PMC7133468/](https://pmc/articles/PMC7133468/)
103. Smith H. Clinical AI: opacity, accountability, responsibility and liability. *AI Soc [Internet].* 2020 Jul 25 [cited 2021 Apr 26];1:3. Available from: <https://doi.org/10.1007/s00146-020-01019-6>



# CHAPTER

2

## MAINSTREAMING PERSONALIZED HEALTHCARE – TRANSFORMING HEALTHCARE THROUGH NEW ERA OF ARTIFICIAL INTELLIGENCE

Ketan Paranjape  
Michiel Schinkel  
Prabath W.B. Nanayakkara

## ABSTRACT

### Background

Medicine has entered the digital era, driven by data from new modalities, especially genomics and imaging, as well as new sources such as wearables and Internet of Things. As we gain a deeper understanding of the disease biology and how diseases affect an individual, we are developing targeted therapies to personalize treatments. There is a need for technologies such as Artificial Intelligence (AI) to be able to support predictions for personalized treatments. In order to mainstream AI in healthcare we will need to address issues such as explainability, liability and privacy. Developing explainable algorithms and including AI training in medical education are many of the solutions that can help alleviate these concerns.

### Methods

A systematic literature search was performed searching the electronic database PubMed from inception to December 2019.

### Results

Presented these findings at the IEEE Biomedical and Health Informatics Conference.

### Conclusion

Both personalized healthcare and AI are evolving. As we understand more about the biology, diagnostics, and augment medical knowledge with patient data from images, genomics, and medical records, we will be able to identify personalized therapies for individuals. As we gain a deeper understanding of how AI works, healthcare professionals will be able to explain the decision they make with the help of AI tools. With the help of technology and regulatory bodies we will be able to resolve challenges with liability and privacy. We are well on our way to provide personalized treatment strategies driven by AI.

### Keywords

Artificial intelligence, Deep learning, Machine learning, Personalized healthcare

## Highlights

1. 21<sup>st</sup> century healthcare professionals are confronted by many technological advancements and large amounts of data including the digitization of healthcare due to government regulations (e.g. HITECH Act)
2. Along with data from traditional sources such as medical records, health data is now available from real-world evidence, molecular information (genomics), wearables, mobile applications, payers and clinical trials.
3. Technologies like artificial intelligence can help us analyze these vast amounts of data to derive insights and help with clinical decision making.
4. AI has shown great promise in treating certain disease (lung cancer) and still needs work in others (Sepsis).
5. Challenges with AI such as explainability and liability need to be resolved.

## INTRODUCTION

21<sup>st</sup> century healthcare professionals are confronted by many technological advancements and large amounts of data. Physicians and nurses are overwhelmed by data from infusion pumps, vital sign monitors, laboratory tests, molecular tests, medical images and all the data that has been recorded in electronic medical records (1) (2). Gathering this data and using it to make an informed and personalized decision poses a unique challenge that has yet to be overcome. New technologies such as artificial intelligence (AI) have the intrinsic ability to gain insights from large amounts of data from various sources and may be used to solve these problems.

## DIGITIZATION OF HEALTHCARE

An explosion of data and knowledge in medicine, diseases and science is beginning to impact the healthcare industry, bringing with it a real transformation in care. The Health Information Technology for Economic and Clinical Health Act (HITECH Act) of 2009 resulted in Electronic Health Record (EHR) adoption to increase from 9.4% in 2008 to 83.8% in 2015 (3) through financial incentives and increased penalties for violation of the HIPAA privacy and security rules (4).

Along with digitized medical records, it is estimated that by 2020, medical knowledge will double itself every 73 days (5). A doctor would need to spend 29 hours a day absorbing new medical knowledge to stay up to date. In other words, we have reached the capacity of the human brain and time to follow and process the new medical knowledge that is being generated and published.

In an era of digital technology, we will be able to increasingly tailor medical treatment to the needs of individuals and small groups of patients. More information will be captured, stored and analyzed to learn how diseases manifest themselves and how patients experience them every day. Combined with a deeper understanding of molecular science and new methods for diagnostics, this development will bring disruptive change to how we research, develop, approve and pay for medicines, as well as how patients and their physicians make decisions about whether, when and how to treat their illnesses.

## NEW SOURCES OF DATA

As knowledge in medicine, diseases and science grows, high-quality data from a wide array of sources can be collected for each patient and can be connected to data from large pools of other patients for analysis (6) (7). This enables us to arrive at a deeper understanding of disease biology and its expression in individual patients (8). Patients are more knowledgeable and informed, and in the position to demand innovative and effective treatments. Real-world evidence (9), molecular information generated from next-generation sequencing (10) (11), data from wearable devices (12) and mobile apps (13) and novel clinical trials (14) (15) are increasing our understanding of health and disease. The regulatory environment needs to and is evolving and adjusting for these novel approaches to healthcare (16). The task of unlocking the ecosystem of digital healthcare

cannot be done by anyone alone. As a result, new types of partnerships are forming to ensure we are moving towards value-based, personalized patient care (17) (18).

## ARTIFICIAL INTELLIGENCE IN HEALTHCARE

With the digitization of healthcare, technologies such as AI can help us analyze these vast amounts of data to derive insights and help with decision making.

AI in healthcare is the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data without direct human input. Since a seminal paper by Sir Alan Turing in 1950 (19), AI has had many advances in Natural Language Processing (NLP) (20), Machine Learning (21), Deep Learning (22), Speech Recognition (23), Virtual Agents (24), and AI-optimized Hardware (25), amongst others.

Today, AI is already used in healthcare (26) for example to decrease false-positive results in screening for breast cancer (27) (28), reduce medical transcription costs (29), improve physician workflow while relieving and helping to prevent burnout (30), robotic surgery resulting in shorter length of hospitalization and loss of blood (31) and predicting mortality rates of patients with acute heart failure (32).

In the past, the most important stakeholder in healthcare, which is the patient, suffered from a broad category of diseases which were treated with the same medicines, leaving physicians to puzzle over why they worked for some people and not others.

Today scientists have begun to understand, target, and diagnose illnesses on an individual level and AI can play a significant role in this process given its unique capabilities of detecting subtle disease specific patterns from a wide array of sources, such as molecular diagnostics, that humans would never recognize.

## PERSONALIZED MEDICINE

With the use of machine learning applications, a subcategory of AI, that can combine data from all state-of-the-art diagnostic tests and other resources, there is more potential for personalized medicine than ever before. A high-level discussion of two specific fields of medicine will show what AI, in combination with all these new technologies, can and cannot do.

### Lung Cancer

A 2018 narrative review on AI applications for non-small cell lung cancer shows that there are already many applications being tested in this field (33). Machine learning algorithms can be used to increase our understanding of important genomic pathways in lung cancer, with the use of microarray data (34). Also, machine learning can be used to predict which patient will respond to newly developed checkpoint inhibitors (35) or personalize radiation therapy (36), thereby choosing an optimal treatment strategy. A key feature in the success of AI for lung cancer is that many molecular abnormalities have already been discovered, such as mutations in the epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) (33). These very specific markers provide an excellent starting point for algorithms to work from.

## Sepsis

A similar narrative review of AI applications for sepsis was published in 2019 (37), showing that applications to improve diagnosis, treatment and prognosis exist already. Many algorithms to predict sepsis onset have been developed, with encouraging results (38) (39). However, there are no clear molecular abnormalities on which new algorithms can be trained. The rapid onset and heterogeneous presentation of this syndrome makes it so, that the understanding of pathophysiology remains poor when compared to that of lung cancer. The potential of AI is therefore limited, as unique features needed to do adequate predictions are not yet known. Machine learning has the ability to classify in the absence of unique features, but to detect conditions like sepsis more data is needed because of heterogeneous presentation and unique features that are needed in order to provide understanding to develop new treatments.

Algorithms can be trained to predict the best possible treatment on an individual level (40), but can only consider the general treatments that exist today - antibiotics, source control and intravenous fluids. Likely, better treatment options exist, but the machine learning algorithms are limited by human knowledge at this point in time.

For AI to be able to provide personalized predictions for treatment, meaningful data at scale is needed. Clinical trial data, molecular data and general patient data needs to be integrated in advanced predictive models. A broad understanding of pathophysiology in a certain field is needed in order for AI to become valuable.

## MAINSTREAMING ARTIFICIAL INTELLIGENCE IN HEALTHCARE

As discussed, some disease specific challenges, such as with sepsis, hold back the mainstream adoption of AI in certain fields for now, but there are also some general concerns and challenges about the adoption of AI in healthcare which have to be addressed at a larger scale.

### Challenges with AI

Challenges with the introduction of AI in healthcare are centered around explainability (41), liability (42) (43) and privacy. Furthermore, the medical educational system for healthcare professionals will have to undergo a rigorous transformation.

Lack of explainability of AI algorithms is likely to bring about some resistance by the medical community. The more accurate the algorithms, such as neural networks, the less explainable they are. This “black box” phenomenon (43) makes it hard for healthcare professionals to get used to working with AI and trusting the algorithm (44). In the end, physicians still have to make a final decision and not knowing why you would make a certain decision will raise many more issues when a patient is given the wrong diagnosis. Software developers will have to take this into account and prioritize both explainability and accuracy. Having explainability will likely also simplify acceptance by the US Food and Drug Administration (FDA) as mentioned in the recent documents regulating AI (45) and presumably by medical regulatory agencies around the world.

Then there is an issue with liability. Who is to blame when something goes wrong? There is no case law about the use of medical AI yet. Even worse, the current laws seem to “incentivize physicians to minimize the potential value of AI” (42) as they will only face liability when current protocols are not adhered to. New malpractice laws will have to be developed to specify the liability of all involved parties: healthcare professionals, hospitals, software companies, software developers and the data collectors.

Privacy is another outstanding issue with the use of AI. Vast amounts of patient data are needed for some AI algorithms to properly function. Google for example is using 46 billion data points collected from 216,221 adults’ de-identified data over 11 combined years from two hospitals to predict the outcomes of hospitalized patients (46) (47). This raises questions about how this data is obtained and whether all patients have had a fair chance to decide about the use of their data.

Lastly, as patients begin to see the benefits of AI and proactively use tools like the chatbot (48), physicians will need to be aware of limitations of such technologies and care for the patient accordingly (49). They will need to be trained in how to effectively use such technologies to their benefit and help ease their burden (50).

## Resolving Challenges

In order to alleviate the main concern with explainability we need models that can explain the why, so a physician can confidently diagnose a patient with a certain disease. Explainable AI is a new emerging discipline that is working towards making machine decisions transparent, interpretable, traceable, and reproducible (51) (52) (53).

The healthcare community needs to be educated regarding these challenges and how to address them and also establish standards and guidelines so a physician and a machine working together has the greatest potential to improve clinical decision-making and patient health outcomes.

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 (54).

In Jun 2018, the American Medical Association’s House of Delegates comprised of proportional representations of every major national medical specialty society and state medical associations adopted its first policy on healthcare Augmented Intelligence (55). Some of the recommendations included identifying opportunities to integrate the perspective of practicing physicians into the development, design, validation and implementation of healthcare AI; encouraging education for patients, physicians, medical students, other healthcare professionals, and health administrators to promote greater understanding of the promise and limitations of healthcare AI; and exploring the legal implications of healthcare AI, such as issues of liability or intellectual property, and advocate for appropriate professional and governmental oversight for safe, effective, and equitable use of and access to healthcare AI.

## CONCLUSIONS

Both personalized healthcare and AI are evolving. As we understand more about the biology, diagnostics, and augment medical knowledge with patient data from images, genomics, and medical records, we will be able to identify personalized therapies for individuals. As we gain a deeper understanding of how AI works, healthcare professionals will be able to explain the decision they make with the help of AI tools. With the help of technology and regulatory bodies we will be able to resolve challenges with liability and privacy. We are well on our way to provide personalized treatment strategies driven by AI.

## REFERENCES

1. J. A. Hazelzet, "Can fuzzy logic make things more clear?," *Crit. Care*, vol. 13, no. 1, p. 116, 2009.
2. "Big Data in Healthcare: Challenges & Promise | NEJM Catalyst." [Online]. Available: <https://catalyst.nejm.org/big-data-healthcare/>. [Accessed: 26-Nov-2019].
3. "Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015." [Online]. Available: <https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php>. [Accessed: 25-Nov-2019].
4. J. Adler-Milstein, A. J. Holmgren, P. Kralovec, C. Worzala, T. Searcy, and V. Patel, "Electronic health record adoption in US hospitals: the emergence of a digital 'advanced use' divide," *J. Am. Med. Informatics Assoc.*, vol. 24, no. 6, pp. 1142–1148, Nov. 2017.
5. P. Densen, "Challenges and Opportunities Facing Medical Education," *Trans. Am. Clin. Climatol. Assoc.*, vol. 122, p. 48, 2011.
6. P. Deshpande *et al.*, "Big Data Integration Case Study for Radiology Data Sources," in *2018 IEEE Life Sciences Conference (LSC)*, 2018, pp. 195–198.
7. S. Tsumoto and S. Hirano, "Healthcare IT: Integration of consumer healthcare data and electronic medical records for chronic disease management," in *2014 IEEE International Conference on Granular Computing (GrC)*, 2014, pp. 310–315.
8. M. R. S. Rao and President, "Disease biology in the coming decade(s).," *Ann. Neurosci.*, vol. 19, no. 4, pp. 149–50, Oct. 2012.
9. V. B. Katkade, K. N. Sanders, and K. H. Zou, "Real world data: an opportunity to supplement existing evidence for the use of long-established medicines in health care decision making.,," *J. Multidiscip. Healthc.*, vol. 11, pp. 295–304, 2018.
10. S. Yohe and B. Thyagarajan, "Review of Clinical Next-Generation Sequencing," *Arch. Pathol. Lab. Med.*, vol. 141, no. 11, pp. 1544–1557, Nov. 2017.
11. M. Zhao, D. Liu, and H. Qu, "Systematic review of next-generation sequencing simulators: computational tools, features and perspectives," *Brief. Funct. Genomics*, vol. 16, no. 3, p. elw012, Apr. 2016.
12. G. Gresham *et al.*, "Wearable activity monitors in oncology trials: Current use of an emerging technology," *Contemp. Clin. Trials*, vol. 64, pp. 13–21, Jan. 2018.
13. M. Sharp and D. O'Sullivan, "Mobile Medical Apps and mHealth Devices: A Framework to Build Medical Apps and mHealth Devices in an Ethical Manner to Promote Safer Use - A Literature Review.,," *Stud. Health Technol. Inform.*, vol. 235, pp. 363–367, 2017.
14. E. R. Dorsey, C. Venuto, V. Venkataraman, D. A. Harris, and K. Kieburtz, "Novel Methods and Technologies for 21st-Century Clinical Trials," *JAMA Neurol.*, vol. 72, no. 5, p. 582, May 2015.
15. E. L. Korn and B. Freidlin, "Adaptive Clinical Trials: Advantages and Disadvantages of Various Adaptive Design Elements," *JNCI J. Natl. Cancer Inst.*, vol. 109, no. 6, Jun. 2017.
16. "Software as a Medical Device (SaMD) | FDA." [Online]. Available: <https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd>. [Accessed: 25-Nov-2019].
17. G. S. Ginsburg and K. A. Phillips, "Precision Medicine: From Science To Value," *Health Aff.*, vol. 37, no. 5, pp. 694–701, May 2018.
18. K. A. Teng and D. L. Longworth, "Personalized healthcare in the era of value-based healthcare," *Per. Med.*, vol. 10, no. 3, pp. 285–293, May 2013.
19. "Computing Machinery and Intelligence on JSTOR." [Online]. Available: [https://www-jstor-org.iclbezp1.cc.ic.ac.uk/stable/2251299?seq=1#metadata\\_info\\_tab\\_contents](https://www-jstor-org.iclbezp1.cc.ic.ac.uk/stable/2251299?seq=1#metadata_info_tab_contents). [Accessed: 28-Jul-2019].

20. K. Kreimeyer *et al.*, "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," *J. Biomed. Inform.*, vol. 73, pp. 14–29, Sep. 2017.
21. B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine Learning for Medical Imaging," *RadioGraphics*, vol. 37, no. 2, pp. 505–515, Mar. 2017.
22. R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.
23. B. Gold, N. Morgan, and D. Ellis, *Speech and Audio Signal Processing*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.
24. "The Future of Customer Service Is AI-Human Collaboration." [Online]. Available: <https://sloanreview.mit.edu/article/the-future-of-customer-service-is-ai-human-collaboration/>. [Accessed: 28-Jul-2019].
25. "Machine Learning Benchmark Set with IBM POWER9 and GPUs." [Online]. Available: <https://www.ibm.com/blogs/research/2018/03/machine-learning-benchmark/>. [Accessed: 28-Jul-2019].
26. K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nat. Biomed. Eng.*, vol. 2, no. 10, pp. 719–731, Oct. 2018.
27. J.-J. Mordang, A. Gubern-Mérida, A. Bria, F. Tortorella, G. den Heeten, and N. Karssemeijer, "Improving computer-aided detection assistance in breast cancer screening by removal of obviously false-positive findings," *Med. Phys.*, vol. 44, no. 4, pp. 1390–1401, Apr. 2017.
28. K. J. Dreyer and J. R. Geis, "When Machines Think: Radiology's Next Frontier," *Radiology*, vol. 285, no. 3, pp. 713–718, Dec. 2017.
29. K. Saxena, R. Diamond, R. F. Conant, T. H. Mitchell, I. G. Gallopy, and K. E. Yakimow, "Provider Adoption of Speech Recognition and its Impact on Satisfaction, Documentation Quality, Efficiency, and Cost in an Inpatient EHR.," *AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci.*, vol. 2017, pp. 186–195, 2018.
30. "3 ways medical AI can improve workflow for physicians | American Medical Association." [Online]. Available: <https://www.ama-assn.org/practice-management/digital/3-ways-medical-ai-can-improve-workflow-physicians>. [Accessed: 03-Feb-2019].
31. Medical Advisory Secretariat, "Robotic-assisted minimally invasive surgery for gynecologic and urologic oncology: an evidence-based analysis.," *Ont. Health Technol. Assess. Ser.*, vol. 10, no. 27, pp. 1–118, 2010.
32. J.-M. Kwon *et al.*, "Artificial intelligence algorithm for predicting mortality of patients with acute heart failure.," *PLoS One*, vol. 14, no. 7, p. e0219302, 2019.
33. M. Rabbani, J. Kanevsky, K. Kafi, F. Chandelier, and F. J. Giles, "Role of artificial intelligence in the care of patients with nonsmall cell lung cancer," *European Journal of Clinical Investigation*, vol. 48, no. 4. Blackwell Publishing Ltd, 01-Apr-2018.
34. T. Kikuchi *et al.*, "Expression profiles of non-small cell lung cancers on cDNA microarrays: Identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs," *Oncogene*, vol. 22, no. 14, pp. 2192–2205, Apr. 2003.
35. P. Charoentong *et al.*, "Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade," *Cell Rep.*, vol. 18, no. 1, pp. 248–262, Jan. 2017.
36. G. Amit *et al.*, "Automatic learning-based beam angle selection for thoracic IMRT," *Med. Phys.*, vol. 42, no. 4, Apr. 2015.
37. M. Schinkel, K. Paranjape, R. S. N. Panday, N. Skyttberg, and P. W. B. Nanayakkara, "Clinical applications of artificial intelligence in sepsis: A narrative review," *Comput. Biol. Med.*, p. 103488, Oct. 2019.
38. R. J. Delahanty, J. Alvarez, L. M. Flynn, R. L. Sherwin, and S. S. Jones, "Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis.," *Ann. Emerg. Med.*, 2019.

39. Q. Mao *et al.*, "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU," *BMJ Open*, vol. 8, no. 1, p. e017833, 2018.
40. M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," *Nat. Med.*, vol. 24, no. 11, pp. 1716–1720, 2018.
41. A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?," Dec. 2017.
42. W. N. Price, S. Gerke, and I. G. Cohen, "Potential Liability for Physicians Using Artificial Intelligence," *JAMA*, Oct. 2019.
43. "Artificial Intelligence, Medical Malpractice, and the End of Defensive Medicine | Bill of Health." [Online]. Available: <https://blogs.harvard.edu/billofhealth/2017/01/26/artificial-intelligence-medical-malpractice-and-the-end-of-defensive-medicine/>. [Accessed: 14-Oct-2018].
44. R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," *Sci. Rep.*, vol. 6, no. 1, p. 26094, May 2016.
45. "Artificial Intelligence and Machine Learning in Software as a Medical Device | FDA." [Online]. Available: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>. [Accessed: 09-Dec-2019].
46. "Google is using 46 billion data points to predict a hospital patient's future — Quartz." [Online]. Available: <https://qz.com/1189730/google-is-using-46-billion-data-points-to-predict-the-medical-outcomes-of-hospital-patients/>. [Accessed: 01-Nov-2018].
47. A. Rajkomar *et al.*, "Scalable and accurate deep learning for electronic health records," Jan. 2018.
48. M. Bates, "Health Care Chatbots Are Here to Help," *IEEE Pulse*, vol. 10, no. 3, pp. 12–14, May 2019.
49. A. Palanica, P. Flaschner, A. Thommandram, M. Li, and Y. Fossat, "Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey," *J. Med. Internet Res.*, vol. 21, no. 4, p. e12887, Apr. 2019.
50. "How Chatbots Can Ease Primary Care Burden | Medpage Today." [Online]. Available: <https://www.medpagetoday.com/primarycare/generalprimarycare/72802>. [Accessed: 10-Aug-2019].
51. L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagel, "Explaining Explanations: An Overview of Interpretability of Machine Learning," May 2018.
52. M. Wu, M. C. Hughes, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi-Velez, "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability," Nov. 2017.
53. "Interpreting Deep Neural Networks – ShallowMind." [Online]. Available: <http://www.shallowmind.co/jekyll/pixyll/2017/12/30/tree-regularization/>. [Accessed: 26-Nov-2019].
54. K. Paranjape, M. Schinkel, R. Nannan Panday, J. Car, and P. Nanayakkara, "Introducing Artificial Intelligence Training in Medical Education (Preprint)," *JMIR Med. Educ.*, Sep. 2019.
55. "AMA Passes First Policy Recommendations on Augmented Intelligence | American Medical Association." [Online]. Available: <https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence>. [Accessed: 15-Oct-2018].



# CHAPTER

# 3

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

Michiel Schinkel\*

Ketan Paranjape\*

Rishi S. Nannan Panday

Niclas Skytberg

Prabath W.B. Nanayakkara

\* Both authors contributed equally to the manuscript

## ABSTRACT

### Background

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.

### Methods

A systematic literature search was performed searching the electronic database 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.

### Results

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.

### Conclusion

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.

### Keywords

Artificial intelligence, Sepsis, Machine learning, PROBAST, Mortality

### Highlights

1. Artificial Intelligence has potential to improve identification of septic patients.
2. Generalizability of AI algorithms is still poor.
3. AI models are often at high risk of bias due to predictor variables in the outcome.
4. Insufficient availability of data will decrease AI accuracy in clinical practice.
5. There is a large gap between creation and clinical implementation of algorithms.

## 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  $\geq 2$  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.

## 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

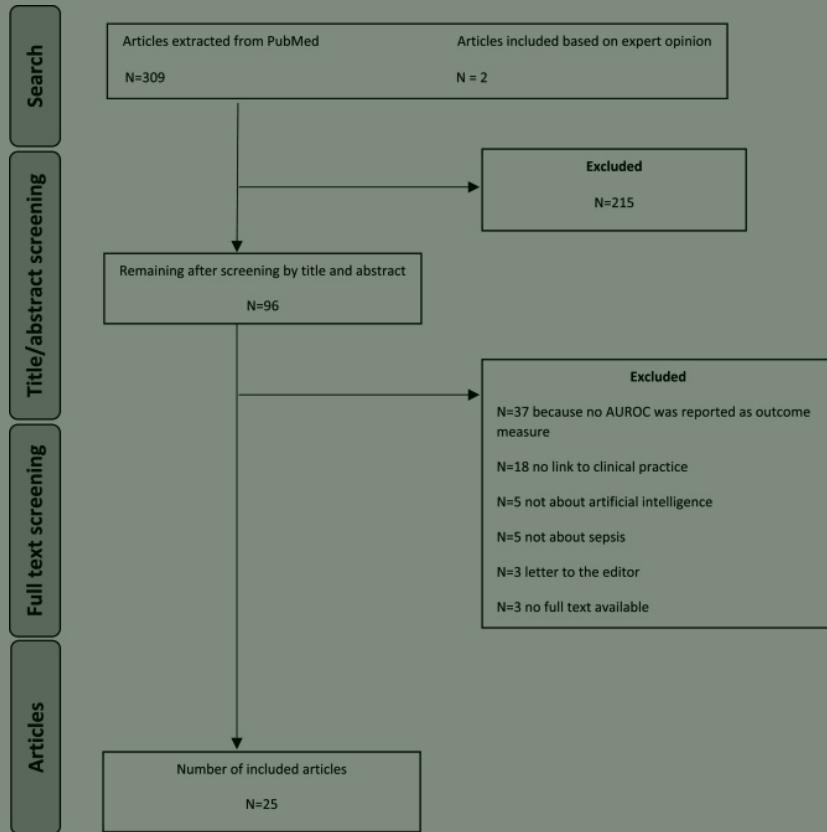
The search, supplemented with two articles based on expert opinion, yielded 311 articles. After screening by title and abstract, ninety-six were selected, as they reported on some application of AI in patients with sepsis. After full texts were appraised, twenty-two articles (10, 23-42) on diagnosis 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



**Figure 1.** Flowchart of article selection

in an intensive care unit (ICU) population (24). 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 (27). 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 (26). Kaji and colleagues predicted same-day and next-day sepsis (25). 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 (28). Saqib et al. reported on a model that predicted sepsis in an ICU population with an AUROC of 0.696 (31). Shashikumar et al. predicted sepsis in an ICU population 4 hours in advance with an AUROC of 0.78 (28). Taneja et al. predicted sepsis onset with a model based on vital parameters, as well as individual biomarkers (29). 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 (30).

Four articles reported on predictions regarding the pathogens that caused sepsis. Van Steenkiste and colleagues used an AI model to predict positive blood cultures (33).

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 (36). 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 (34). Lamping and colleagues used an AI model to distinguish sepsis from non-infectious SIRS in critically ill children, achieving an AUROC of 0.78 (35).

## 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 (41). Furthermore, Aushev et al, reported on a model predicting in-hospital mortality with an AUROC of 0.845 (37). Meiring et al. used an algorithm to predict mortality over time in the ICU (40). The model reached an AUROC of 0.895. The article by Jaimes et al. described a model that predicted 28-day mortality (39). 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 (38). Ward and colleagues predicted 30-day mortality for patients with an infection or sepsis, reaching an AUROC of 0.79 (42).

## 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  $\mu\text{g}/\text{kg}$  versus 1.4  $\mu\text{g}/\text{kg}$ ;  $p < 0.01$ ). 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

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

<b>Author, year</b>	<b>Study design</b>	<b>Setting</b>	<b>Database (MIMIC = Medical Information Mart for Intensive Care)</b>
<b>Diagnosis</b>			
Delahanty, 2019	Retrospective	Emergency Department	Hospital database (2.759.529 patient encounters)
Desautels, 2016	Retrospective	Intensive Care	MIMIC-III
Kaji, 2019	Retrospective	Intensive Care	MIMIC-III
Kam, 2017	Retrospective	Intensive Care	MIMIC-III
Mao, 2018	Retrospective	Hospital wide	Hospital database (17.467.987 patient encounters) MIMIC-III
Nemati, 2017	Retrospective	Intensive Care	Hospital database (27.527 patient encounters) MIMIC-III
Taneja, 2017	Retrospective	Hospital wide	Hospital database (444 patient encounters)
Henry, 2015	Retrospective	Intensive care	MIMIC-III
Saqib, 2018	Retrospective	Intensive care	MIMIC-III
Shashikumar, 2017	Retrospective	Intensive Care	Hospital database (242 patient encounters)
Barton, 2019	Retrospective	Hospital Wide	Hospital database (91,445 patient encounters) MIMIC-III
<b>Pathogen prediction</b>			
Van Steenkiste, 2018	Retrospective	Hospital wide	Hospital database (2177 patient encounters)
Oonsivalai, 2018	Retrospective	Hospital wide	Hospital database (243 patient encounters)
Lamping, 2018	Prospective, RCT	Pediatric ICU	Hospital based (230 patient encounters)
Ratzinger, 2018	Prospective	Hospital wide	Hospital based (466 patient encounters)
<b>Prognosis</b>			
Aushev, 2018	Retrospective	Intensive care	ShockOmics
Dybowski, 1996	Retrospective	Intensive care	Hospital database (4484 patient encounters)
Garcia-Gallo, 2018	Retrospective	Intensive care	MIMIC-III
Jaimes, 2005	Retrospective	Emergency department	Hospital database (542 patient encounters)
Meiring, 2018	Retrospective	Intensive care	MIMIC-II MIMIC-III
Taylor, 2016	Retrospective	Emergency department	Hospital database (4676 patient encounters)
Ward, 2017	Retrospective	Trials/Studies	Hospital database (2514 patient encounters)
<b>Treatment assistance</b>			
Komorowski, 2018	off-policy evaluation	Intensive care	MIMIC-III eICU
Merouani, 2008	Prospective, randomized	Intensive care	Hospital database (42 patient encounters)
Shimbukuro, 2017	Randomized controlled trial	Intensive care	Hospital database (142 patient encounters)

No. predictor variables in model	Outcome	PROBAST-assessment (Risk of bias; concern with applicability)
13	AUROC: 0.93 at 1-hour, AUROC 0.97 at 24-hours	ROB: high, applicability: high
8	AUROC: 0.880 at disease onset	ROB: high, applicability: low
119	AUROC: 0.952 at same-day, 0.876 at next-day	ROB: high, applicability: unclear
9	AUROC: 0.929	ROB: high, applicability: low
6	AUROC: 0.92 4-hours before sepsis onset.	ROB: high, applicability: low
65	AUROC: 0.85 4-hours before sepsis	ROB: high, applicability: high
21	AUROC: 0.81 at disease onset	ROB: high , applicability: high
26	AUROC: 0.83 28.2-hours before sepsis onset.	ROB: high , applicability: high
12	AUROC: 0.696	ROB: high, applicability: low
Unclear	AUROC: 0.78 4-hours before sepsis onset.	ROB: high, applicability: low
6	AUROC: 0.83 48-hours before onset.	ROB: high, applicability: low
9	AUROC: 0.99 with 72 hours of data	ROB: low, applicability: low
35	AUROC: 0.80 for ceftriaxone susceptibility	ROB: high, applicability: high
8	AUROC: 0.78 for infectious vs. non-infectious SIRS	ROB: high, applicability: high
21	AUROC: 0.73 for bacteraemia.	ROB: high, applicability: low
80	AUROC: 0.845 for ICU mortality	ROB: high, applicability: high
11	AUROC: 0.863 for in-hospital mortality	ROB: high, applicability: high
18	AUROC: 0.8083 for 1 year mortality	ROB: high, applicability: low
10	AUROC: 0.8782 for 28-day mortality	ROB: low, applicability: low
25	AUROC: 0.895 for mortality at ICU discharge	ROB: low, applicability: low
25	AUROC: 0.86 for in-hospital mortality	ROB: high, applicability: high
18	AUROC: 0.79 for 30-day mortality	ROB: high, applicability: high
48	AI policy associated with lowest mortality	ROB: high, applicability: high
2	Median duration of shock significantly shorter (28.5 hours versus 57.5 hours).	ROB: -, applicability: -
8	In-hospital mortality decreased by 12.4 percentage points	ROB: high, applicability: low

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

Study	Risk of bias (ROB)			Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB
<b>Diagnosis</b>								
Dehahanty, 2019	+	-	-	+	+	+	-	-
Desautels, 2016	-	-	-	-	+	+	-	+
Kajii, 2019	?	-	-	-	?	+	-	?
Kam, 2017	+	-	-	-	+	+	-	+
Mao, 2018	-	-	-	+	+	+	-	+
Nemati, 2017	-	-	-	-	-	-	-	-
Taneja, 2011	-	-	-	-	-	-	-	-
Henry, 2015	+	-	-	-	-	-	-	-
Saqib, 2018	-	-	+	+	-	-	+	-
Shashikumar, 2017	+	+	-	-	+	+	-	+
Barton, 2019	+	+	-	+	+	+	-	+
<b>Pathogen prediction</b>								
Van Steenkiste, 2018	+	+	+	+	+	+	+	+
Oonsivilai, 2018	+	-	+	-	-	-	-	-
Lamping, 2018	+	+	+	-	+	+	-	-
Ratzinger, 2018	+	+	+	-	+	+	-	+
<b>Prognosis</b>								
Aushhev, 2018	-	-	+	-	+	-	+	-
Dybowski, 1996	-	+	+	-	-	+	+	-
Garcia-Gallo, 2018	-	+	+	+	+	+	-	+
Jaimes, 2005	+	+	+	+	+	+	+	+

**Table 2.** (continued)

Study	Risk of bias (ROB)			Applicability				Overall
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	
<b>Treatment assistance</b>								
Morabowski, 2018	-	-	+	+	+	+	+	+
Taylor, 2016	+	+	+	+	+	-	+	-
Ward, 2017	-	-	+	-	+	-	+	-
<b>PROBAST = Prediction model Risk Of Bias ASSESSment Tool; ROB = risk of bias. * + indicates low ROB/low concern regarding applicability; - indicates unclear ROB/undear concern regarding applicability; and ? indicates high ROB/high concern regarding applicability.</b>								

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

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

**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

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 (27). 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-centre 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 favourable 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 database 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 and with algorithms that used streams of physiologic data to detect sepsis early. We agree that the use of these routinely measured biomarkers 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.

## REFERENCE LIST

1. Purkayastha S, Gichoya JW, Siva ;, Addeppally A. Implementation of a single sign-on system between practice, research and learning systems. 2017;
2. Hazelzet JA. Can fuzzy logic make things more clear? Crit Care. 2009;13(1):116.
3. Bali J, Garg R, Bali R. Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? Vol. 67, Indian Journal of Ophthalmology. Wolters Kluwer Medknow Publications; 2019. p. 3–6.
4. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health. 2018;8(2).
5. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng [Internet]. 2018 Oct 10 [cited 2019 Jul 28];2(10):719–31. Available from: <http://www.nature.com/articles/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 [Internet]. 2018 Nov 27 [cited 2019 Apr 2];19(6):1236–46. Available from: <https://academic.oup.com/bib/article/19/6/1236/3800524>
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 [cited 2019 Apr 2]. p. 97–106. Available from: [http://link.springer.com/10.1007/978-3-642-33612-6\\_11](http://link.springer.com/10.1007/978-3-642-33612-6_11)
8. Gottesman O, Johansson F, Komorowski M, Faisal A, Sontag D, Doshi-Velez F, et al. Guidelines for reinforcement learning in healthcare. Nat Med [Internet]. 2019 Jan 7 [cited 2019 Apr 2];25(1):16–8. Available from: <http://www.nature.com/articles/s41591-018-0310-5>
9. Clermont G. Artificial neural networks as prediction tools in the critically ill. Crit Care. 2005;9(2):153–4.
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):1146–50.
11. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign. Crit Care Med. 2013 Feb;41(2):580–637.
12. 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. 2016 Feb;315(8):801.
13. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019 May;321(20):2003.
14. Quinten VM, van Meurs M, Ligtenberg JJ, Ter Maaten JC. Prehospital antibiotics for sepsis: beyond mortality? Lancet Respir Med. 2018 Mar;6(3):e8.
15. Goulden R, Hoyle M-C, Monis J, Railton D, Riley V, Martin P, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. Emerg Med J. 2018 Jun;35(6):345–9.
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 Apr;0(0).
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 Oct;25(5):328–34.

18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec;316(22):2402.
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* [Internet]. 2017 Feb 25 [cited 2019 Jul 28];542(7639):115–8. Available from: <http://www.nature.com/articles/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 Mar;8(1):109–18.
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):3042–54.
22. Moons KGM, Wolff RF, Riley RD, Penny ;, Whiting F, Westwood M, 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.
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;
24. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, 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.
25. Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. *PLoS One*. 2019;14(2):e0211057.
26. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med*. 2017;89:248–55.
27. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, 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.
28. Shashikumar SP, Stanley MD, Sadiq I, Li Q, Holder A, Clifford GD, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol*. 2017;50(6):739–43.
29. Taneja I, Reddy B, Damhorst G, Dave Zhao S, Hassan U, Price Z, et al. Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis. *Sci Rep*. 2017;7(1):10800.
30. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* [Internet]. 2015;7(299):299ra122. Available from: <http://stm.sciencemag.org/content/7/299/299ra122.full%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26246167>
31. 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;2018:4038–41.
32. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, 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 Jun;109:79–84.
33. Van Steenkiste T, Ruyssinck J, De Baets L, Decruyenaere J, De Turck F, Ongenae F, et al. Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. *Artif Intell Med*. 2018;
34. Oonsivilai M, Mo Y, Luangasanatip N, Lubell Y, Miliya T, Tan P, 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.

35. Lampert F, Jack T, Rübsamen N, Sasse M, Beerbaum P, Mikolajczyk RT, 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.
36. Ratzinger F, Haslacher H, Perkmann T, Pinzan M, Anner P, Makristathis A, 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.
37. Aushev A, Ripoll VR, Vellido A, Aletti F, Pinto BB, Herpain 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.
38. 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;
39. Jaimes F, Farbizar 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.
40. Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, et al. Optimal intensive care outcome prediction over time using machine learning. *PLoS One.* 2018;13(11):e0206862.
41. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, 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):269–78.
42. 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.
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):1716–20.
44. Merouani M, Guignard B, Vincent F, Borron SW, Karoubi P, Fosse J-P, et al. Norepinephrine weaning in septic shock patients by closed loop control based on fuzzy logic. *Crit Care.* 2008;12(6):R155.
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.
46. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA.* 2017 Oct;318(13):1241.
47. A computational approach to early sepsis detection. *Comput Biol Med.* 2016 Jul;74:69–73.
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.
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):547–53.
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.
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 May;13(5):426–35.
52. Mauri C, Principe L, Bracco S, Meroni E, Corbo N, Pini B, 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.

53. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* [Internet]. 2016 Dec 24 [cited 2019 Jul 19];3(1):160035. Available from: <http://www.nature.com/articles/sdata201635>
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):907–10.
55. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. *AMIA . Annu Symp proceedings AMIA Symp* [Internet]. 2013 [cited 2019 Mar 10];2013:1109–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24551396>
56. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* [Internet]. 2012 Feb [cited 2019 Mar 10];3(1):9–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23236300>
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;
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):e495–503.
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.
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 May;23(3):978–86.





# 4

## CHAPTER

### TOWARDS UNDERSTANDING THE EFFECTIVE USE OF ANTIBIOTICS FOR SEPSIS

Michiel Schinkel\*

Ketan Paranjape\*

Justin Kundert

Rishi S. Nannan Panday

Nadia Alam

Prabath W.B. Nanayakkara

\* Both authors contributed equally to the manuscript

*Chest. 2021 Oct;160(4):1211-21*

## 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

## 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 treatments5. 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  $\geq 38^{\circ}\text{C}$  or  $\leq 36^{\circ}\text{C}$ . 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

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

<b>Analysis Trial Number</b>	<b>K-Centroids Method</b>	<b>Min/Max Cluster Parameters</b>	<b>Number of Traits Evaluated</b>	<b>Traits Assessed</b>	<b>Number of Clusters for Partitioning (Based on Preliminary Diagnostics Assessment)</b>
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; Heart Rate (Ambulance); Temperature (Ambulance)	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
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)	2
14	K-means	2/8	4	Respiratory Rate (Ambulance); Delta Respiratory Rate (Ambulance > ED); Heart Rate (Ambulance > ED)	2
15	K-medians	2/8	4	Respiratory Rate (Ambulance); Delta Respiratory Rate (Ambulance > ED); Heart Rate (Ambulance > ED)	2
16	K-means	2/8	3	C-reactive Protein (ED Lab); Leucocytes (ED Lab); Creatinine (ED Lab)	4

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

<b>Analysis Trial Number</b>	<b>K-Centroids Method</b>	<b>Min/Max Cluster Parameters</b>	<b>Number of Traits Evaluated</b>	<b>Traits Assessed</b>	<b>Number of Clusters for Partitioning (Based on Preliminary Diagnostics Assessment)</b>
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

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

K-Centroids Method	Min/Max Number of Traits	Cluster Parameter Evaluated	Number of Traits Assessed	Diagnostic Results					Cluster Results				
				Number of Clusters (k) for Partitioning	Adjusted Rand (Mean)	Calinski-Harabasz (Mean)	Cluster	Size	Average Distance Separation		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		
K-means	2/8	2	Heart Rate (Ambulance); Temperature (Ambulance)	5	0.80	5266.8	1	734	5.92	13.5	8.87		
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		

## 4

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).

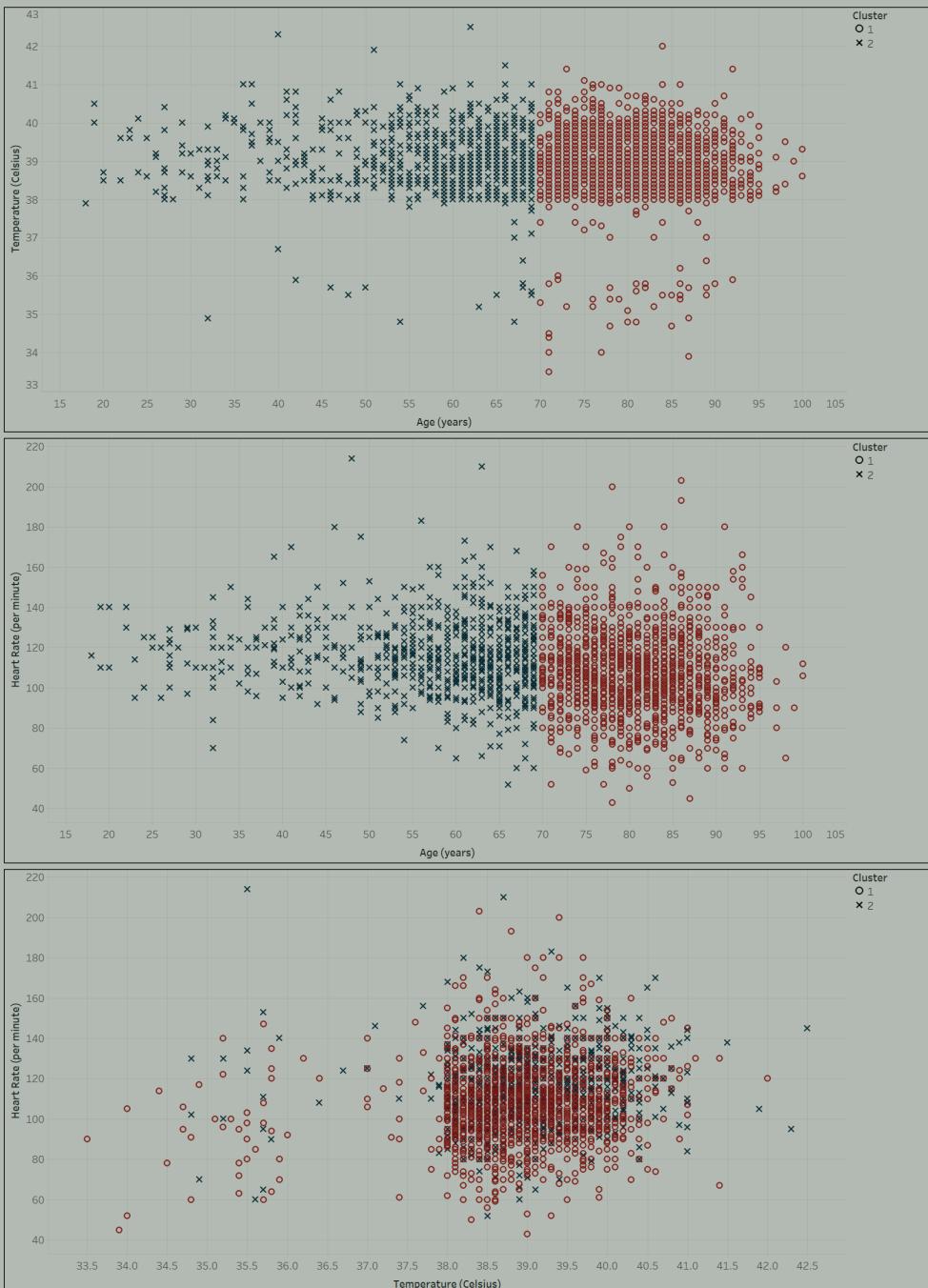
In some cases, age was not used as a continuous variable, but as a dichotomous variable. The categories were created by splitting the dataset in the 50% youngest and 50% oldest patients, in order to obtain equally large numbers of patients in both groups (22). The age ranges in these groups were 18 -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

## RESULTS

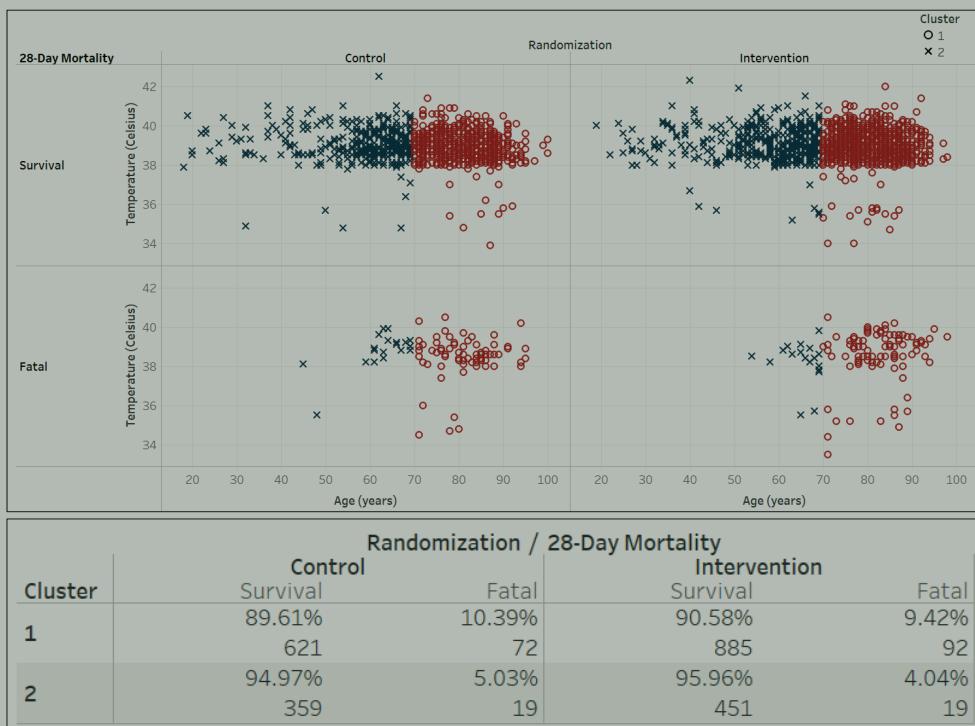
### Exploratory partitioning cluster analysis

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.

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



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



**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.

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 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 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 ( $OR = 1.03$ ;  $95\%-CI = 1.00-1.05$ ;

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

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

$p = 0.066$ ) 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

**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

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 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. e-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;  $p = 0.03$ ).

### 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<sup>24-26</sup>. 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

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

Characteristics	Age continuous				Age dichotomous			
	Raw model		Adjusted model		Raw		Adjusted model	
Intervention (Y)	0.13 (0.02-1.10)	0.061 (0.01-0.79)	0.031 (0.01-0.80)	0.07 (0.01-0.80)	0.031 (0.02-1.10)	0.68 (0.02-1.10)	0.126 (0.36-1.06)	0.082 (0.34-1.03)
Age	1.03 (1.01-1.05)	0.001 (1.01-1.05)	0.008 (0.99-1.03)	1.00 (0.99-1.03)	0.583 (1.14-2.77)	1.77 (1.00-2.59)	0.012 (0.54-1.51)	0.053 (0.54-1.51)
Age * intervention	1.03 (1.00-1.05)	0.066 (1.00-1.06)	0.033 (1.00-1.07)	1.03 (1.00-1.07)	0.030 (0.90-3.05)	1.65 (0.99-3.63)	0.110 (0.99-3.63)	0.055 (1.11-4.30)
Sex (F)				0.91 (0.66-1.24)	0.543 (1.10-1.26)			0.92 (0.67-1.26)
Charlson comorbidity index (per point increase)				1.17 (1.09-1.25)	0.001 (1.04-1.20)	1.12 (1.04-1.20)	0.002 (1.10-1.26)	<0.001 (1.04-1.20)
qSOFA (lower than 2)				0.46 (0.33-0.63)	0.001 (0.40-0.78)	0.56 (0.33-0.62)	<0.001 (0.33-0.62)	0.45 (0.33-0.62)
Do not resuscitate order (Y)					3.75 (2.58-5.55)	<0.001 (2.58-5.55)		4.17 (2.88-6.14)
Antibiotics prior to hospital visit (Y)					1.34 (0.93-1.91)	0.111 (0.93-1.91)		1.32 (0.91-1.88)
Immunosuppressive conedication (Y)					1.48 (1.00-2.16)	0.046 (1.00-2.16)		1.46 (0.98-2.13)
Positive blood culture (Y)					1.37 (0.95-1.96)	0.088 (0.95-1.96)		1.38 (0.95-1.97)
Ceftriaxone resistant pathogen (Y)					2.83 (0.38-14.00)	0.235 (0.38-14.00)		2.55 (0.33-13.35)

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<sup>27</sup>. Recent research indicates that early administration of antibiotics is associated with higher mortality when given to patients with greater diagnostic uncertainty<sup>28</sup>. 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.

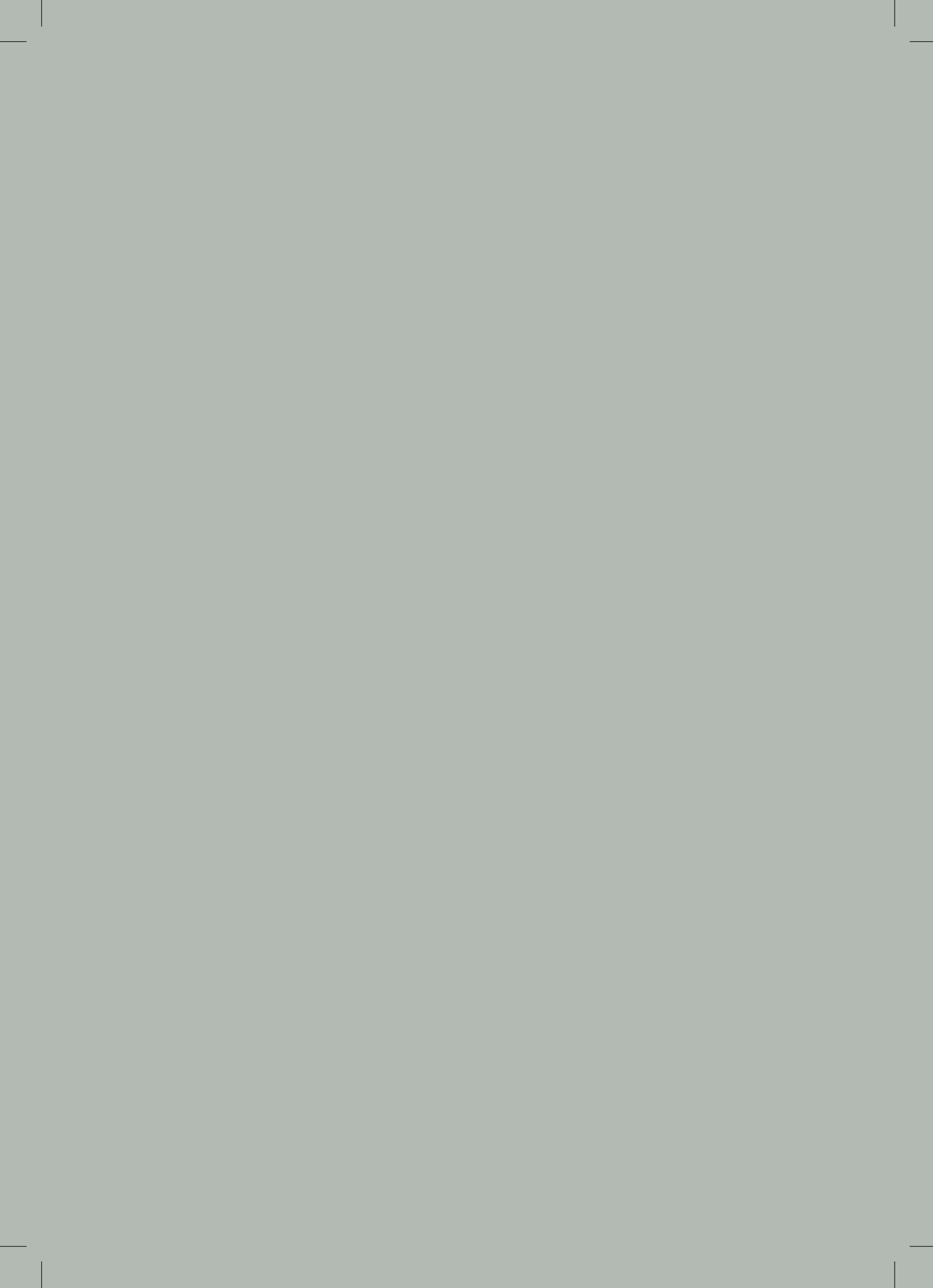
## FINANCIAL SUPPORT

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 Groot 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, Barnecke 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 Vugt 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-Lloeches 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](http://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





# 5

## CHAPTER

### DEVELOPMENT OF A 'META-MODEL' TO ADDRESS MISSING DATA, PREDICT PATIENT-SPECIFIC CANCER SURVIVAL AND PROVIDE A FOUNDATION FOR CLINICAL DECISION SUPPORT

Jason M. Baron

Ketan Paranjape

Tara Love

Vishakha Sharma

Denise Heaney

Mathew Prime

## ABSTRACT

### Objective

Like most real world data, EHR-derived data from oncology patients typically exhibits wide inter-patient variability in terms of available data elements. This inter-patient variability leads to missing data and can present critical challenges in developing and implementing predictive models to underlie clinical decision support (CDS) for patient-specific oncology care. Here we sought to develop a novel ensemble approach to addressing missing data that we term the "meta-model" and apply the meta-model to patient-specific cancer prognosis.

### Methods

Using real-world data, we developed a suite of individual random survival forest models to predict survival in patients with advanced lung, colorectal cancer (CRC) and breast cancer. Individual models varied by the predictor data used. We combined models for each cancer type into a meta-model that predicted survival for each patient using a weighted-mean of the individual models for which the patient had all requisite predictors.

### Results

The meta-model significantly outperformed many of the individual models and performed similarly to the best performing individual models. Comparisons of the meta-model to a more traditional imputation-based method of addressing missing data supported the meta-model's utility.

### Conclusions

We developed a novel machine learning-based strategy to underlie CDS and predict survival in cancer patients, despite missing data. The meta-model may more generally provide a tool for addressing missing data across a variety of clinical prediction problems. Moreover, the meta-model may address other challenges in clinical predictive modeling including model extensibility and integration of predictive algorithms trained across different institutions and datasets.

**Key Words:** Missing Data, Imputation, Clinical Decision Support, Meta-model, Machine Learning, Survival

## BACKGROUND AND SIGNIFICANCE

Predictive models trained using real world clinical data offer tremendous potential to provide patients and their clinicians patient-specific information regarding diagnosis, prognosis or optimal therapeutic course.(1-10) For example, a recent high-profile study trained a machine learning model using hundreds of thousands of patient records to forecast the development of acute kidney injury.(9) However, key challenges have limited the introduction of machine learning-based predictive models into real clinical settings. (3) One set of challenges relates to inter-patient variability in data availability. In most real world datasets, many patients will lack recorded findings for many clinical factors. (3 7 11-13) For example, some hospitals may have a laboratory test menu that includes more than 1,000 unique orderable tests. Most patients will have had at most a small fraction of these possible tests. A similar pattern involving substantial "missing data" would usually be observed for non-laboratory clinical data, including other diagnostic studies, elements of patient history and physical exam findings. The issue of data heterogeneity becomes particularly significant when considering time series data; even patients who have similar diagnostic tests or physical exam maneuvers performed may have them at different time points or repeated at varying intervals.(3)

Many commonly used machine learning algorithms require complete data sets and cannot directly use for training or prediction datasets containing missing data. Data scientists commonly employ several strategies to enable use of real world data that include missing elements in predictive analyses. One strategy involves pre-processing a set of clinical data by "imputing" missing data elements. (14) While there are numerous variations on imputation and related approaches including single imputation, multiple imputation, and expectation maximization, most imputation approaches are fundamentally designed to use available data to estimate the distribution or value of each element of missing data. (7 11-13 15-18) The pre-processed dataset, including both actual and imputed clinical findings, can then be used to train standard machine learning models or can be applied to trained models to generate predictions. However, imputation, while very useful in many contexts has important limitations. Most imputation algorithms assume data are "missing at random" (MAR); since diagnostic studies are selected and ordered in response to the clinical setting, most clinical datasets will violate the MAR assumption.(3 11 18) Likewise, imputation can introduce additional uncertainty and inaccuracy into predictions and may obscure some of the intuition behind some predictive models.

As described below, we propose and demonstrate an alternative approach to imputation in addressing missing data. We term this new approach, the "meta-model". To develop and apply the meta-model, we consider the problem of patient-specific prognosis prediction in patients with advanced oncologic disease. While population-based survival statistics are available across a wide range of cancer types and patients, patient-specific information can be harder to discern. For example, based on national SEER statistics, the overall five-year survival of patients with stage IV colon cancer is just 14%. (19) However, some individual patients will have a considerably better than average survival. The critical question for an oncologist then, when seeing an individual

patient, is not the population survival but what the individual patient's prognosis is. Individualizing patient prognosis is not itself a new endeavor. On the contrary, numerous published studies describe clinical risk factors that portend better or worse prognosis. For example, prior studies clearly establish that patients with colon cancer experience shorter survival on average if they have comorbid diabetes.(20) While a clinician may take these types of published findings into account when considering prognosis, their true clinical utility can be quite limited. In particular, patients may have multiple clinical factors that individually could convey improved or worsened prognosis; there would usually not be a viable strategy to calculate the aggregate impact of these multiple factors. Indeed, prior studies have shown limitations of the human brain in manually making predictions based on a large number of predictors.(2) Thus, as a secondary focus of this manuscript, we propose, validate and demonstrate a strategy to apply machine learning to the development of patient-specific Kaplan-Meier survival curves. These patient-specific curves may offer oncologists and other clinicians the opportunity to more accurately assess patient-prognosis and communicate risk to patients.

## OBJECTIVES

This manuscript has two objectives. The primary objective is to develop and demonstrate a novel "meta-model" approach to addressing missing data. As described in detail below, our meta-model concept includes an ensemble of underlying models based on varying predictors with the final output based on an aggregate of all individual models for which a patient has complete data. The meta-model may also address other challenges in predictive CDS implementation including model extensibility and integration of predictive algorithms trained across different institutions and datasets.

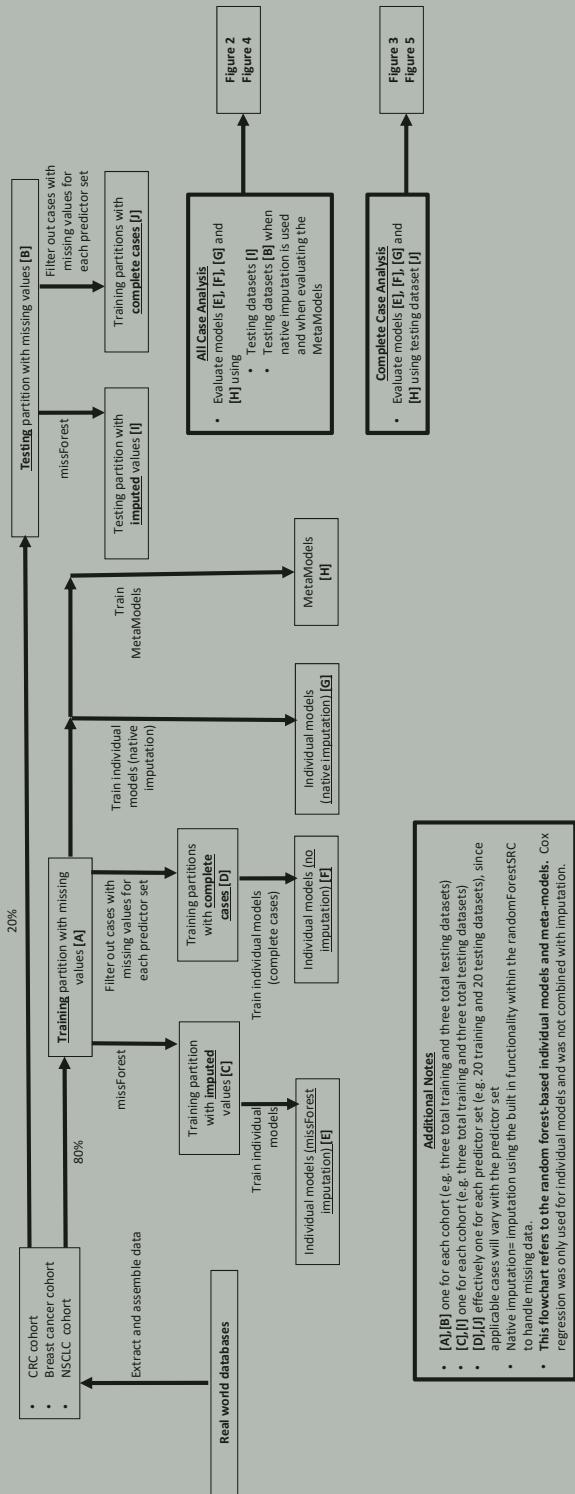
The secondary objective of this manuscript is to develop a method for generating patient-specific Kaplan-Meier survival curves.

## METHODS

An overview of our methods is shown as **FIGURE 1**. Using clinical data from patients with metastatic colorectal cancer, metastatic breast cancer and advanced lung cancer, we first developed a set of survival prediction models intended to individualize patient prognosis. After developing and validating the individual models, we combined the individual models for each dataset into a "meta-model." We demonstrate that this meta-model can provide an alternative to imputation in addressing missing data and may offer several key advantages. Key methodologic points are described below with additional detail provided in the Supplemental Methods.

### Patient Cohorts

We defined patient cohorts from three subsets of the nationwide Flatiron Health electronic health record-derived de-identified database (21): i) Metastatic CRC (colorectal cancer); ii) Advanced NSCLC (non-small cell lung cancer) iii) Metastatic Breast Cancer. For each patient



**FIGURE 1.** Overview of Data, Model Development and Performance Assessment. Shown is an overview of the approach used to extract, analyze and model the data and assess model performance. Details are described in the Methods section and in the Supplemental Methods.

in our cohort, we extracted and assembled outcome data (time surviving after the date of advanced diagnosis) and selected clinical features that were commonly available and that we thought might help to predict prognosis ("predictors"). We randomly split the cases for each cohort into a training and a testing partition in an approximately 80:20 ratio.

## Survival Outcome

We captured survival, defined as the number of days between advanced tumor diagnosis and death for use as our outcome variable. Patients were censored to the time of their last encounter.

## Predictors and Predictor Sets

We considered patient demographics, tumor characteristics, molecular biomarkers and laboratory test results for use as predictors (**TABLE 1**). We prioritized potential predictors for inclusion based on factors including data availability and expected predictive value. That is, we favored predictors available on a larger number of patients (based on a preliminary exploration of the data) and those, which, based on our domain expertise, we thought were more likely to be of value. In addition, we considered the usability of predictors both in our analysis and in potential downstream applications, (e.g. we preferred predictors often available in structured form with meanings that are substantially standardized across institutions).

We further grouped the predictors into "predictor sets" (**TABLE 1**). Similar to how we selected the predictors themselves, we selected the predictor set groupings based on patterns of data availability (e.g. we preferentially included lab tests commonly performed together in the same predictor set) with an emphasis on developing predictor sets for which many or all patients would have all of the available predictors. The number of total cases along with an accounting of censored patients and deceased patients, available for use with each predictor set, is shown as **TABLE 2**.

## Individual Model development

For each cohort and set of predictors, we trained two different survival models: one linear Cox regression model and another based on a random forest. We used the R "survival" package (22) to develop and validate the linear models and the R "randomForestSRC" package (23-26) to develop the tree-based models. We included 100 trees per individual model. Additional detail on these models is available in the **Supplemental Methods**.

## Addressing Missing Data in Training and Testing Individual Models

We used three strategies to train and test individual models in the setting of missing data. The first strategy involved using complete cases with respect to each predictor set (i.e. patients were excluded from training and/or testing who did not have all of the requisite predictors needed). The second strategy involved imputation. We imputed missing predictor values using the random forest- based imputation algorithm,

**TABLE 1.** Predictors used in each model

Variable	Type	Advanced Lung Cancer								Metastatic Breast Cancer					Metastatic Colorectal Cancer							
		A	B	C	D	E	F	G	H	A	B	C	D	E	A	B	C	D	E	F	G	
Gender	Categorical	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1
GroupStage	Categorical	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1
dxageYrs	Numerical	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1
CrcSite	Categorical	x	x	x	x	x	x	x	x	x	x	x	x	x	0	1	1	1	1	1	1	1
Histology	Categorical	0	1	1	1	0	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x
SmokingStatus	Categorical	0	1	1	1	0	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x
leukocytes	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
hemoglobin	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
platelets	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
hematocrit	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
erythrocytes	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
creatinine	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
lymphocytes	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	1
protein	Lab	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0	0	0
bilirubin	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	1	1	1
urea.nitrogen	Lab	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	1	1	1	1	1
carcinoembryonic.ag	Lab	0	0	0	0	0	0	0	1	x	x	x	x	x	0	0	0	0	1	1	1	1
calcium	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	0	0	0
sodium	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	0	0	0
potassium	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	0	0	0
alkaline.phosphatase	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	0	0	0
lymphocytes.per.100.	Lab	x	x	x	x	x	x	x	x	0	0	0	0	1	0	0	0	1	1	0	0	0
leukocytes	Lab	0	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x
monocytes.per.100.	Lab	0	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x
leukocytes	Lab	0	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x
carbon.dioxide	Lab	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x
monocytes	Lab	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x
chloride	Lab	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x
lactate.dehydrogenase	Lab	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x
ER	Categorical	x	x	x	x	x	x	x	x	0	0	0	1	1	0	0	0	0	0	0	0	0
PR	Categorical	x	x	x	x	x	x	x	x	0	0	0	1	1	0	0	0	0	0	0	0	0
HER2	Categorical	x	x	x	x	x	x	x	x	0	0	0	1	1	0	0	0	0	0	0	0	0
EGFR	Categorical	0	0	0	0	1	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x
ALK	Categorical	0	0	0	0	1	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x
KRAS	Categorical	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	1	1	0
glucose	Lab	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	1	1	0	0	0
BRAF	Categorical	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	1	0
Random number	Numerical	0	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x

1 → predictor is included in specified predictor set; 0 → predictor is NOT included in specified predictor set, but is included in other predictor sets for the cancer type; x → predictor is not included in any predictor sets within the corresponding cancer type

**TABLE 2.** Patient characteristics by predictor set

Cohort	Predictor Set	Complete Cases (N)	Complete cases (percent of cohort)	Total Deceased	Total Censored
Advanced Lung Cancer	A	6558	100%	1146	5412
	B	6559	100%	1146	5413
	C	6558	100%	1146	5412
	D	4479	68%	639	3840
	E	3254	50%	577	2677
	F	3253	50%	577	2676
	G	2285	35%	338	1947
	H	157	2%	21	136
Metastatic Breast Cancer	A	5045	100%	1633	3412
	B	5045	100%	1633	3412
	C	5045	100%	1633	3412
	D	4795	95%	1548	3247
	E	2854	57%	818	2036
Metastatic Colorectal Cancer	A	6742	100%	1760	4982
	B	6743	100%	1761	4982
	C	6742	100%	1760	4982
	D	3888	58%	888	3000
	E	3435	51%	768	2667
	F	2759	41%	588	2171
	G	1163	17%	269	894

missForest as described in greater detail in the **Supplemental Methods**. The third strategy to addressing missing data involved leveraging built in functionality to handle missing data within the randomforestSRC package ((23-26); we term this third strategy “native imputation”.

### Meta-model Conceptual Approach

Aiming to integrate survival predictions across individual models, improve overall prediction accuracy and offer other important practical properties as described subsequently, we developed an approach we termed the “meta-model”. Conceptually, the meta-model (**SUPPLEMENTAL FIGURE S1**) starts by training individual prediction models, such as the individual survival models described above. It then assigns a weight to each model, based on the model’s accuracy, such that models that tend to predict outcomes with greater accuracy are more heavily weighted (specific approach to assigned weights described below). The meta-model can then be applied to test patients by computing all individual models for which the patient has the necessary predictors and then taking a weighted average of the predictions produced by these individual models. We note that the meta-model in part represents an adaptation of Breiman’s stacked regression. (27)

## Development of the Survival Meta-model

We trained one meta-model for each cohort (three meta-models total) using the predictor sets shown in **Table 1**. For each predictor set, we trained a random forest-based model using methods paralleling the development of the individual survival prediction models. (We refer to each of these individual random forest models within the meta-model as an "individual model".) To assign a weight to each individual model, we performed five-fold cross validation of each individual model capturing the median cross validation AUC of the model at 500,1000 and 1500 days. We then transformed these median AUC values into model weights using one of several weighting functions having the form  $w = (x - 0.5)^n$ , where  $w$  is the weight assigned to the model prediction,  $x$  represents the median cross validation AUC and  $n$  represents an exponent. (See **Supplemental Methods** for additional detail). A value of  $n=2$  was used for all analyses unless otherwise specified.

## Model Evaluation Metrics

We evaluated our models by comparing predicted survival to actual survival for patients in the test partition. We primarily used area under the receiver operating characteristic curve (AUC), describing each models ability to discriminate patients alive vs. deceased at various time points. We specifically considered AUC at 500, 1000 and 1500 days post advanced diagnosis. We calculated AUC using the method described by Heagerty et al. (28) as implemented in the R package survivalROC (28 29). In addition, as described in greater detail in the **Supplemental Methods**, we evaluated model calibration by comparing the actual deaths to the predicted deaths for groups of patients over various time windows.

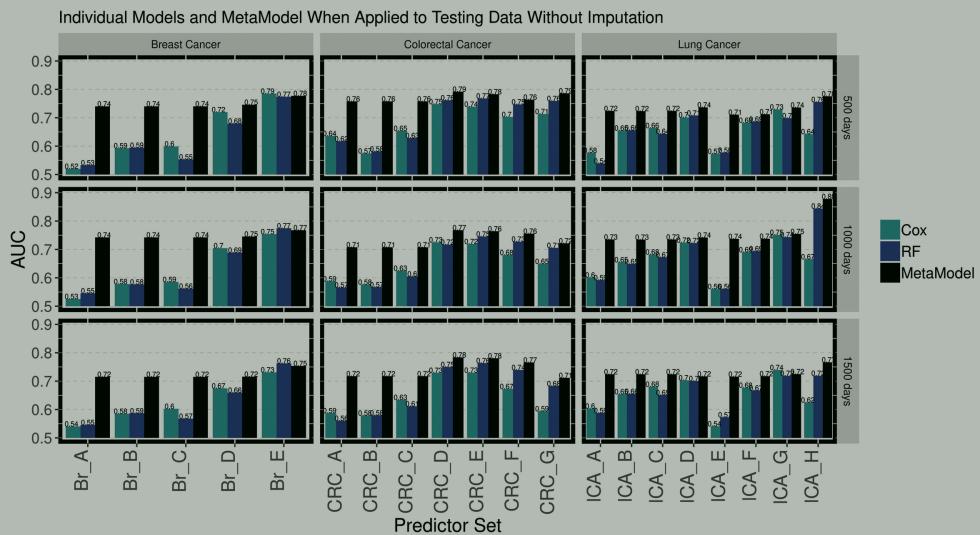
## RESULTS

As described in the methods and **FIGURE 1**, we tested our models in two settings: 1) "Complete Case Analysis" (we used only test cases with all needed predictors available) and 2) "All Case Analysis" (we used test cases regardless of predictor data availability).

### Complete Case Analysis

As shown in **FIGURE 2**, we first considered the performance of individual models in comparison to the meta-model when applied to complete cases within the test data. Although the meta-model is intended for application to patients with a wide range of predictor data availability, for comparison purposes we applied the meta-model in this analysis to the same subsets of test data used for the complete case evaluation of each individual models. Figure 2 also includes corresponding individual Cox regression models for comparison.

The best performing individual models achieved an  $AUC > 0.7$  in predicting mortality at 500,1000 and 1500 days. In almost all cases, the meta-models outperformed the individual models on comparable datasets and in some cases achieved AUCs approaching or exceeding 0.8. The difference in performance between individual models



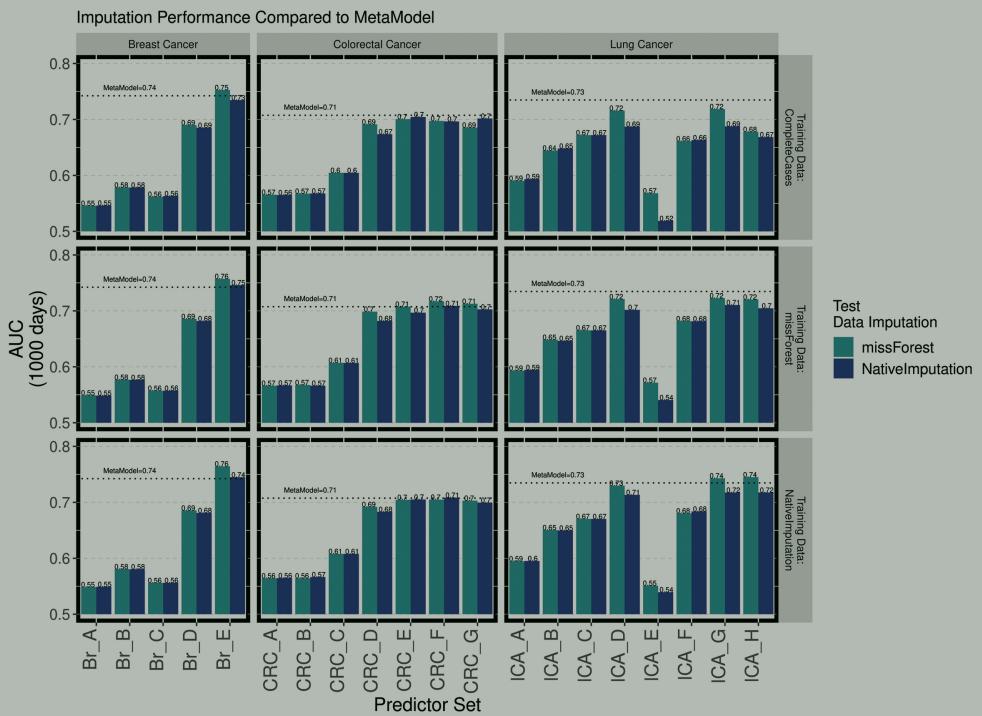
**FIGURE 2.** Model AUC when Applied to Complete Cases from the Test Partition. This figure plots AUC when each model is applied to a subset of patients in the test partition having data available for all of the predictors in the corresponding predictor set. For example, consider the set of bars above “CRC\_D” on the x-axis. The Cox and RF series then denote AUC values for the individual cox and RF models corresponding to CRC predictor set D; the meta-model bar represents the CRC meta-model. All three bars above CRC\_D are based on a subset of the patients in the test partition who have data available for all the predictors included in CRC predictor set D (same subset of test patients used for all three bars). As shown, across all patient subsets, the meta-model performs better than or similar to the individual models. [Cox= individual Cox-regression model; RF=individual random survival forest model; Br\_=breast cancer predictor set; ICA\_=lung cancer predictor set; CRC\_=CRC predictor set].

and the meta-models was most pronounced when considering the simplest individual models (e.g. the “A” and “B” models); this makes sense given that the meta-models in many patients would have been able to leverage a much wider array of predictors than the simple individual models. More interestingly, the meta-models also seem to modestly outperform the more complex individual models in most cases. Supplemental **TABLE S1** provides additional training and testing AUC values.

## All Case Analysis

Since an important goal of the meta-models was applicability to patients with a wide range of predictor data availability, we also compared the meta-model to the individual models when applied to all test patients. For this analysis, we imputed missing test data for application to the individual models. Since the meta-models were designed to accommodate variability in predictor data availability, no imputation was used with the meta-models; however all test patients were used to evaluate both the individual models and the meta-models.

AUC: Model performance in the all case analysis is shown in **FIGURE 3** (for AUCs at 1000 days) and in **SUPPLEMENTAL FIGURES S2 and S3** (for AUCs at 500 and 1500 days,

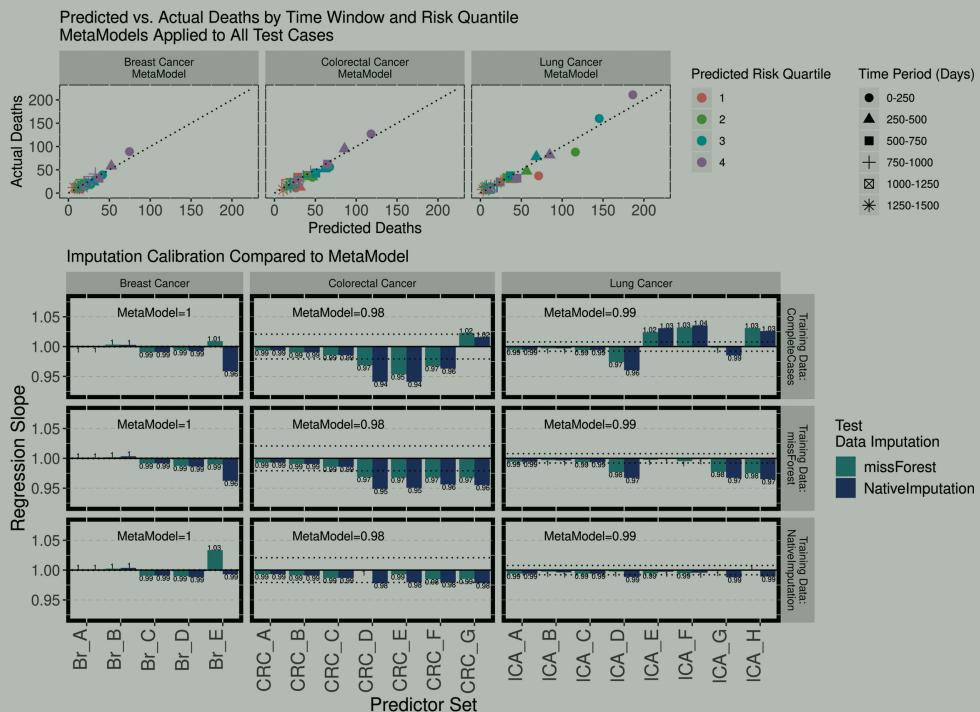


**FIGURE 3.** Model AUC when Applied to All Cases from the Test Partition. Shown is the AUC (1000 days) of each individual model with missing test data addressed using either missForest imputation or imputation within the random survival forest algorithm ("NativeImputation"). For comparison, also shown is the AUC of the corresponding meta-model (dashed line) when applied to the complete set of test patients. All patients within the test partition, regardless of predictor data availability, are included in all analyses. For this analysis, we trained individual models using either complete cases within the training data (top row) or addressed missing training data using missForest (middle row) or native imputation (bottom row). As shown, the meta-model in most cases outperforms the individual models; in some cases, meta-model performance is similar to or negligibly worse than individual models. Analogous figures showing AUC at 500 and at 1500 days are provided as Supplemental **Figure S2** and Supplemental **Figure S3**. [Br\_=breast cancer predictor set; ICA\_=lung cancer predictor set; CRC\_=CRC predictor set].

respectively). As shown, the meta-model significantly outperforms many of the individual models and performs similarly to the best performing individual models. Supplemental **TABLE S1** provides additional training and testing AUC values.

**Impact of Weighting Functions:** We compared various meta-model weighting functions (**SUPPLEMENTAL FIGURE S4**) and found that the specific weighting-function (i.e., value of the exponent n) makes little difference in performance.

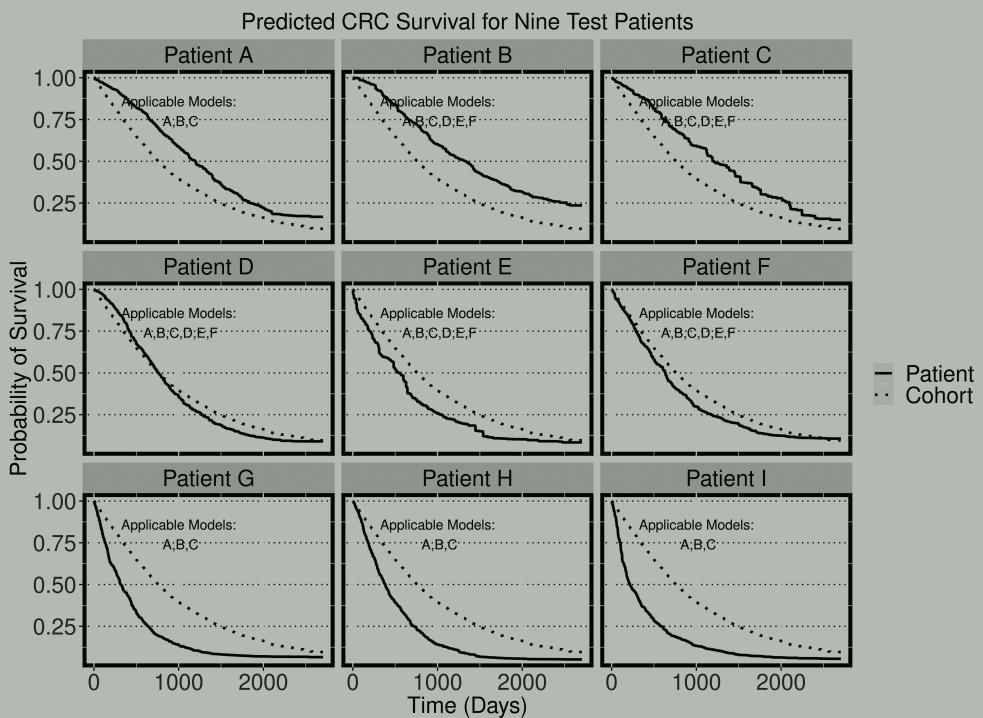
**Model Calibration:** To assess model calibration, we compared predicted mortality to actual mortality for groups of patients over various time windows (**FIGURE 4**). To evaluate further the calibration of each survival model, we fit Poisson regression models with predicted mortality (log transformed) for patient-time groups as the independent variable and actual mortality as the dependent variable. Meta-models produced slopes



**FIGURE 4.** Model Calibration. To test model calibration, we calculated each test patient's predicted mortality over 250 day time windows. We further grouped patients into risk quartiles (4= highest risk of dying; 1= lowest risk) and calculated the aggregate predicted mortality for each patient-group, time-window combination. We then compared predicted mortality to actual mortality. The scatter plots (top) plot actual vs. predicted mortality for each meta-model. The dashed 45-degree line represents perfect calibration. To summarize each model's calibration, we calculated the slope of a Poisson regression line fitting actual-mortality as a function of (log-transformed) predicted mortality with an intercept through the origin. Slopes of one would indicate perfect calibration while slopes substantially different from one would indicate mis-calibration. The bar graphs (bottom) plot the calibration slopes for individual models. The horizontal dashed lines in the graphs represent the meta-model calibration slope (and its mirror image around one). [ Br\_=breast cancer predictor set; ICA\_=lung cancer predictor set; CRC\_=CRC predictor set.]

close to one (0.98-0.99, **FIGURE 4**), suggesting that predicted and actual mortality agree well on average (a slope of 1 would suggest perfect alignment on average). Likewise, the individual models when used with imputation generally showed good calibration, but several cases exhibited slopes considerably further from one (range for individual models =0.94-1.06). **Supplementary Table S2** expands on the analysis in Figure 4 by explicitly considering whether the models are systematically over- or underestimating patient risk in lower- and higher-risk patients.

*Individual Patient Survival:* To illustrate how the meta-model might be used in practice, we plotted individual Kaplan-Meier survival curves for nine selected patients within the test partition of the CRC dataset (**FIGURE 5**).

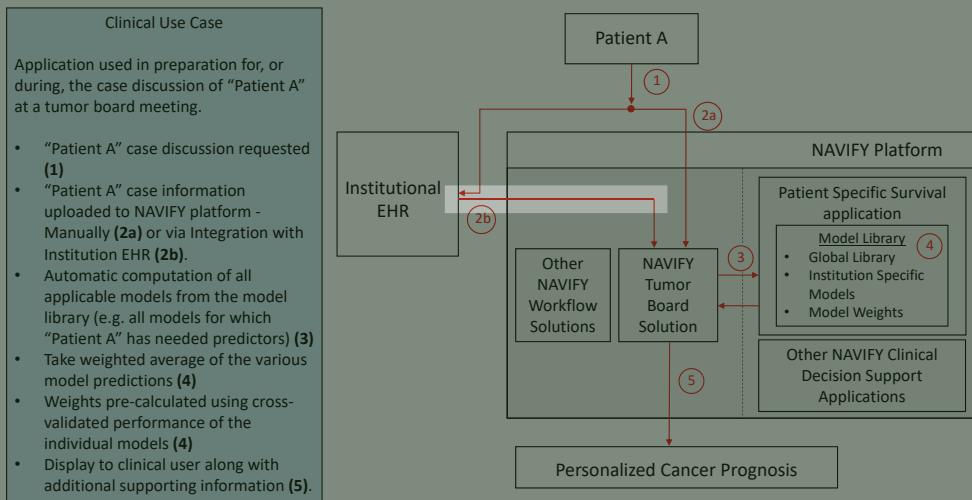


**Figure 5.** Patient-specific Kaplan-Meier Curves for Nine Selected CRC Test Patients. Shown are patient-specific Kaplan-Meier survival curves for nine colorectal cancer patients using meta-model predicted survival probabilities. We selected from the test partition of the CRC data three patients with substantially favorable (top row), three patients with substantially unfavorable (bottom row) and three patients with generally typical (middle row) predicted survival. The solid lines represent the patient's predicted survival and the dashed lines show survival for the cohort as a whole. Applicable models represent the underlying individual CRC models for which the patient had the necessary predictor data and which were included in the prediction.

## DISCUSSION

In this study, we demonstrate the utility both of a meta-model approach to addressing missing data and of the use of machine learning-based models to predict patient survival in advanced colorectal, lung and breast cancer. We show that a meta-model method integrating a suite of underlying models using varied predictors may provide a practical strategy to accommodate missing data. With further validation, we anticipate that the meta-model method described here could be adapted to a wide range of prediction problems, spanning well beyond oncology survival prediction.

The meta-model approach is well suited to the development of clinical decision support, which was our primary aim in undertaking this work. Indeed, as shown in **FIGURE 6**, we aim to build an “app” to provide clinicians with access to patient-specific Kaplan-Meier curves. In addition to addressing missing data, the meta-model may provide a framework for inter-institutional predictive model development. For example, the meta-model



**Figure 6.** A Proposed Schematic of Model Application Infrastructure. Shown is a schematic of an potential infrastructure and workflow within which to implement the meta-model. This is based on the Roche Navify Tumor Board Solution.

could combine individual models trained using separate data sources, even at different institutions. While using multi-institutional data (as opposed to single site data) to train predictive models would often be scientifically desirable in ensuring generalizability and in obtaining data from a sufficient number of patients, administrative challenges to data sharing outside individual health systems often make multi-institutional datasets impractical or impossible to obtain (3). However, the meta-model approach may help to address this challenge to the extent that building a multi-institutional meta-model would only require institutions to share trained underlying individual models and not actual patient data. Moreover, we envision future vendor-developed CDS systems that include both a set of "starter" models as well as functionality for sites to train additional models; the systems could then combine the starter with the locally trained models using the meta-model approach described here. Finally, the meta-model approach may also be useful in identifying tests that were not performed but which could substantially reduce prognostic uncertainty (i.e. tests needed for the better performing underlying models). CDS could recommend the clinician order such tests.

As noted in the introduction, most imputation and other methods for addressing missing data, including the missForest method considered here, assume data are missing at random (MAR). Traditional imputation approaches may be subject to bias when data are not missing at random. For example, consider a hypothetical analysis in which patients with high values for the tumor marker carcinoembryonic antigen (CEA) are more likely to have CEA testing performed. In this case, the "observed" distribution of CEA values (i.e. measured CEA results) will be higher than the unobserved distribution (i.e. what the CEA results would have been in patients who did not have CEA testing). Further, suppose

in this hypothetical example that high CEA correlates with poor prognosis. In this case, imputation might be prone to impute CEA results that are biased high (more in line with the distribution of observed values) and thus a survival prediction algorithm relying on these biased high imputed CEA results might be prone to overly pessimistic prognostic projections in patients without CEA testing. While a formal theoretical evaluation of the extent to which non-random missing data may bias the meta-model is beyond the scope of this manuscript, we postulate that in many cases, the meta-model should be comparatively robust to violations of the MAR assumption. For example, consider what might happen if a meta-model approach were used in the hypothetical CEA scenario noted above. In this case, we might expect individual models that DO include CEA to on average predict poorer prognosis; however, this would be subject to adjustment based on the actual CEA result in these models. Likewise, the individual models that DO NOT include CEA may on average provide overly pessimistic predictions in patients who do not have CEA (this might be similar to the case of imputation) and overly optimistic predictions in those who do. However, these presumably should average out at the population-level if the patterns of missingness in the training and testing partition are the same. Moreover, because models with CEA would be weighted more (if they perform better), patients with CEA may tend to have predictions that overall substantially adjust for CEA results. (This in theory could introduce a net calibration error across the dataset). Future research will be needed to evaluate whether our speculation regarding the robustness of the meta-model to the MAR assumptions holds in practice; however, the calibration data provided here (**FIGURE 4**) empirically may be supportive.

While we had hypothesized that the meta-model would provide better performance than traditional methods of addressing missing data, in our experimental comparison, the meta-model did not universally outperform imputation. When applied to the complete set of test patients and assessed in terms of AUC, imputation in combination with the best performing individual models performed similar to and in some cases, slightly better than the meta-model applied to the same patients. The meta-model may in some cases be more robust to calibration errors introduced due to data missing not at random (see **FIGURE 4** and the paragraph preceding this one). Nonetheless, given the comparable performance of the approach, coupled with improved transparency of the meta-model and the practical applications noted above, we expect that the meta-model will provide a useful tool. Future work will be needed to generalize our assessment of the meta-model to a range of prediction problems. Likewise, we selected only two imputation methods for comparison; we selected missForest in part because it had been shown to work well for laboratory test results in prior research (7 12) and because it can impute both numerical and categorical variables, but additional work comparing the meta-model to additional imputation methods could be informative.

We were surprised that the specific weighting function used to aggregate the predictions from the underlying models had little impact on overall meta-model performance. We had expected that higher values of the exponent  $n$ , which weight better performing underlying models more heavily, would have led to better overall performance. While

we do not have a full explanation, we expect that this may be partly due to the fact that underlying models using differing predictors may provide predictions with at least partially uncorrelated errors. Thus, averaging the "noisy" predictions produced by the underlying models may serve to reduce the overall noise (i.e., overall error). While we selected our predictor set groups manually, in large part based on patterns of data availability, it may be a useful subject of future work to explore characteristics of ideal predictor sets. For example, should correlated predictors be preferentially included in the same or in different predictor sets?

To be sure, our approach is not the first attempt to build patient-specific Kaplan-Meier curves.(30 31) However, most if not all prior attempts to develop patient-specific survival curves have been based on linear models, in contrast to our primary approach. The primary novelty of this manuscript is the use of the meta-model; however, the concept of using patient-specific survival predictions for clinical decision support may also prove to be a useful, practical application of this work.

In addition to the need for future work to further generalize the meta-model approach, the specific application to patient-specific survival prediction is subject to limitations. A key consideration is that in some cases, the algorithms may be providing information that the clinician already knew or suspected; for example, clinicians can of course in some cases use judgement to identify patients who appear sicker and likely have a worse prognosis. While formally testing the clinical value of these algorithms may be a subject of a future study, given the multitude of predictors that went into the algorithms, we hypothesize that it would be difficult for a clinician to manually integrate the value of the many predictors included in our models. Indeed, studies have shown that the human brain is unable to simultaneously integrate a large number of data elements.(2 32) We are considering performing user simulation studies to evaluate how our algorithms perform in comparison to manual clinician intuition.

We are considering several extensions to the patient-specific survival models. In particular, we may develop models for other tumor types and that incorporate additional predictors, including additional biomarkers, co-morbidities, tumor genomics, patient socio-economic factors, care-delivery characteristics and potentially even features extracted from radiologic and whole slide images. Moreover, in addition to providing prognostic predictions, we hypothesize that our approach will be applicable to patient-specific treatment optimization and prescriptive decision support. For example, we plan to explore whether we can update our models to include as predictors the treatment the patient received and then apply counterfactual learning to predict response to therapy.

As we plan our implementation strategy, we will need to carefully evaluate how clinicians and their patients would consider insights provided by our models, and whether such knowledge would have unintended consequences. For example, how would a clinician communicate a personalized life expectancy to a patient, and how will the patient feel about this deeper understanding of their own mortality? Could this information inadvertently bias clinicians when they take treatment decisions to be more or less aggressive than they otherwise might? Who should oversee the appropriateness

of CDS tools for clinical use and monitor for unforeseen outcomes? Addressing these questions may ultimately prove more challenging than the technical aspects of this clinical decision support. (33)

## CONCLUSIONS

The "meta-model" approach we developed and demonstrated in this manuscript offers a strategy to develop clinical predictive models that can accommodate inter-patient heterogeneity in data availability and "missing data". We further demonstrate the value of random forest-based survival models in predicting patient-specific oncology survival. We expect that the proofs of concept we develop here will provide a foundation for novel types of clinical decision support to enable clinicians to make more personalized patient-care decisions.

## ACKNOWLEDGEMENTS

A US patent has been applied for based on this work.

## FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. It was supported using internal Roche Diagnostic funding.

## COMPETING INTERESTS STATEMENT

All authors except Dr. Baron are employed by Roche Diagnostics. Dr. Baron is consultant for Roche Diagnostics. Dr. Prime is also a Director of Open Medical Holdings Ltd, a UK based digital health company.

## CONTRIBUTORSHIP STATEMENT

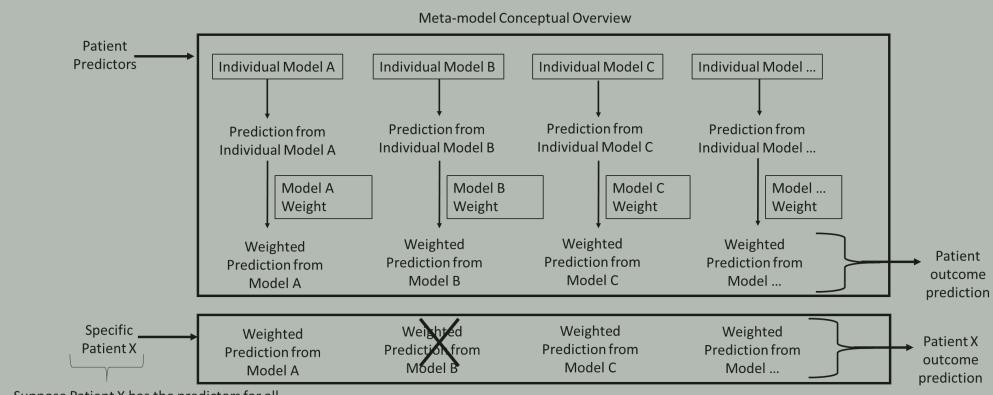
General project framework: JMB, DH, MP; Development of meta-model approach: JMB with input from DH and MP; Model training and validation: JMB; Development of model application: All authors; Manuscript drafting and/or revision: all authors.

## REFERENCES

1. Baron JM, Dighe AS. The role of informatics and decision support in utilization management. *Clin Chim Acta* 2014;427:196-201.
2. Baron JM, Dighe AS, Arnaout R, et al. The 2013 symposium on pathology data integration and clinical decision support and the current state of field. *J Pathol Inform* 2014;5(1):2.
3. Baron JM, Kurant DE, Dighe AS. Machine Learning and Other Emerging Decision Support Tools. *Clin Lab Med* 2019;39(2):319-31.
4. Baron JM, Mermel CH, Lewandrowski KB, Dighe AS. Detection of preanalytic laboratory testing errors using a statistically guided protocol. *Am J Clin Pathol* 2012;138(3):406-13.
5. Kohane IS. HEALTH CARE POLICY. Ten things we have to do to achieve precision medicine. *Science* 2015;349(6243):37-8.
6. Louis DN, Gerber GK, Baron JM, et al. Computational pathology: an emerging definition. *Arch Pathol Lab Med* 2014;138(9):1133-8.
7. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. *Am J Clin Pathol* 2016;145(6):778-88.
8. Rosenbaum MW, Baron JM. Using Machine Learning-Based Multianalyte Delta Checks to Detect Wrong Blood in Tube Errors. *Am J Clin Pathol* 2018;150(6):555-66.
9. Tomasev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572(7767):116-19.
10. Winslow RL, Trayanova N, Geman D, Miller MI. Computational medicine: translating models to clinical care. *Sci Transl Med* 2012;4(158):158rv11.
11. Luo Y, Szolovits P, Dighe AS, Baron JM. 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. *J Am Med Inform Assoc* 2018;25(6):645-53.
12. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* 2013;3(8).
13. Weber GM, Adams WG, Bernstam EV, et al. Biases introduced by filtering electronic health records for patients with "complete data". *J Am Med Inform Assoc* 2017;24(6):1134-41.
14. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and Managing Missing Structured Data in Electronic Health Records: Data Analysis. *JMIR Med Inform* 2018;6(1):e11.
15. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 2007;61(1):79-90.
16. Qi L, Wang YF, He Y. A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Stat Med* 2010;29(25):2592-604.
17. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18(6):681-94.
18. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20(1):40-9.
19. Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) Research Data (1975-2016), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission.
20. Zhu B, Wu X, Wu B, Pei D, Zhang L, Wei L. The relationship between diabetes and colorectal cancer prognosis: A meta-analysis based on the cohort studies. *PLoS One* 2017;12(4):e0176068.

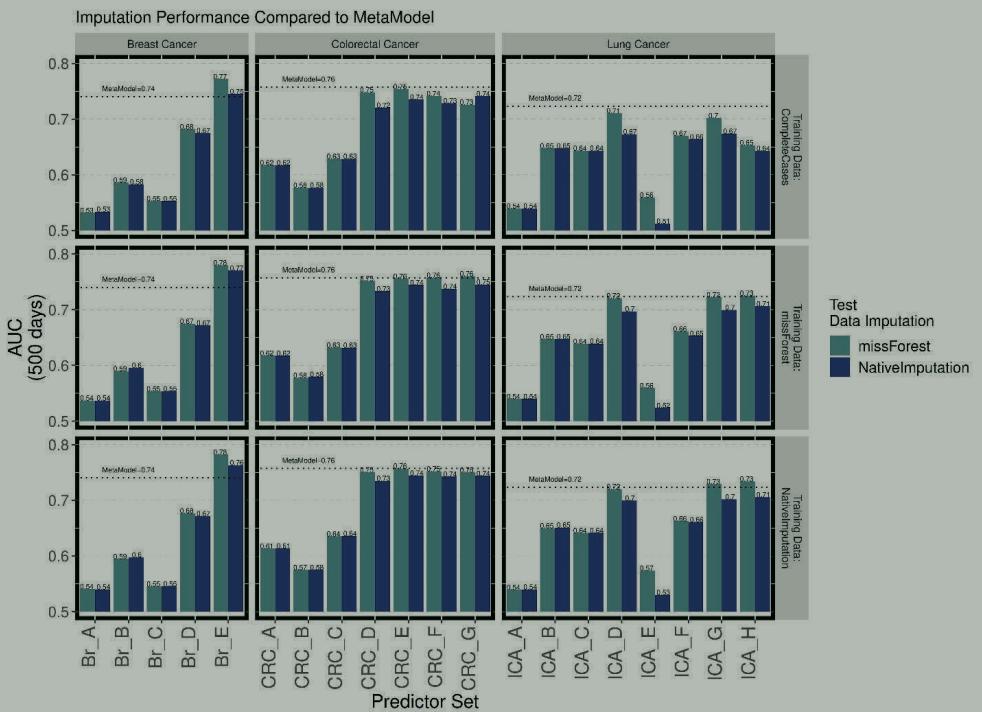
21. Ma X, Long L, Moon S, Adamson BJS, Baxi SS. Comparison of Population Characteristics in Real-World Clinical Oncology Databases in the US: Flatiron Health, SEER, and NPCR. *medRxiv* 2020;2020.03.16.20037143.
22. Therneau TM. A Package for Survival Analysis in S. 2015.
23. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics* 2014;15(4):757-73.
24. Ishwaran H, Kogalur UB. Consistency of random survival forests. *Stat Probabil Lett* 2010;80(13-14):1056-64.
25. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. *Ann Appl Stat* 2008;2(3):841-60.
26. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine* 2019;38(4):558-82.
27. Breiman L. Stacked regressions. *Mach Learn* 1996;24(1):49-64.
28. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000;56(2):337-44.
29. Heagerty PJ. Package 'survivalROC'. Secondary Package 'survivalROC' 2015. <https://cran.r-project.org/web/packages/survivalROC/survivalROC.pdf>.
30. Yu C-NG, Russ; Lin, Hsiu-Chin; Baracos , V. Learning Patient-Specific Cancer Survival Distributions as a Sequence of Dependent Regressors Neural Information Processing Systems (NIPS), 2011.
31. Patient-Specific Survival Prediction (PSSP).
32. Hofman MA. Evolution of the human brain: when bigger is better. *Front Neuroanat* 2014;8:15.
33. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17(1):195.

## SUPPLEMENTAL INFORMATION

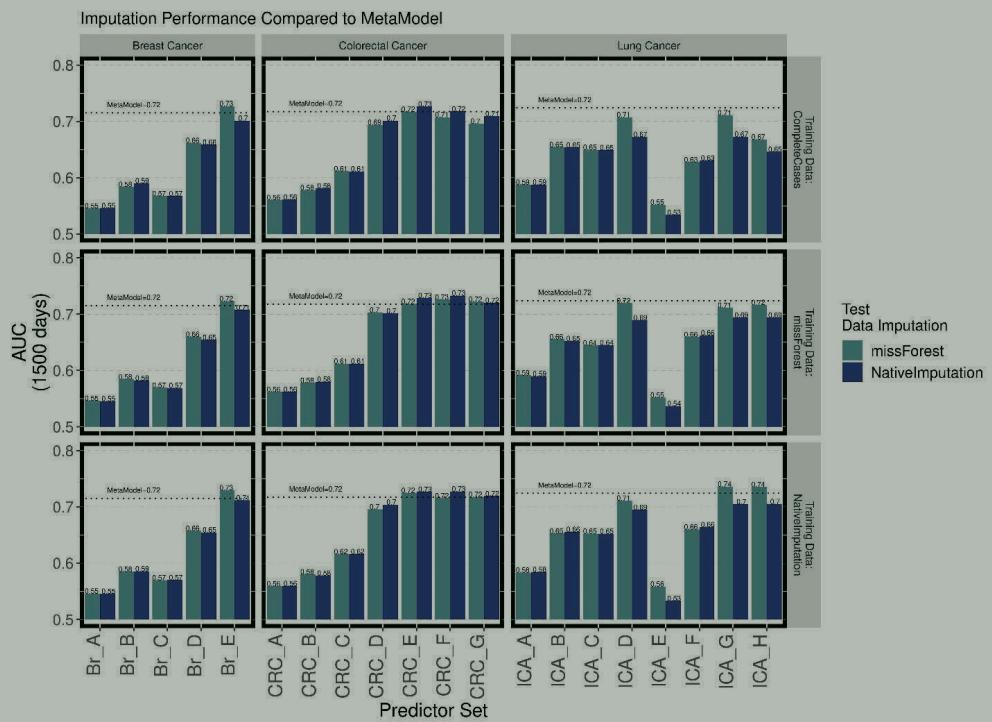


Suppose Patient X has the predictors for all models except model B (e.g. model B requires a molecular biomarker patient X does not have)

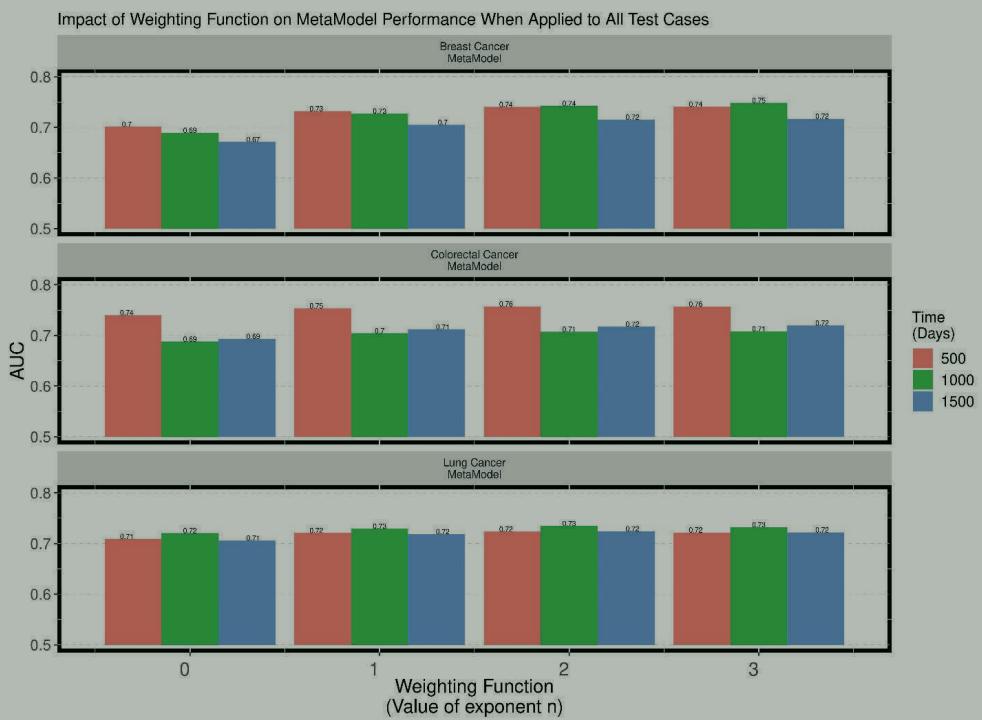
**Supplemental figure S1.** Conceptual Schematic of the Meta-model. Shown is a schematic of a generic meta-model. The upper large box represents a trained meta-model. The meta-model includes an ensemble of individual models trained using complete cases from a set of training data. When applied to subsequent patients, the meta-model computes all individual models for which the patient has the needed predictor data; the final meta-model output represents a weighted average over the applicable individual models. For example, suppose "Patient X" is missing a predictor needed for individual Model B. In this case, the meta-model would make the prediction for patient X using the individual models other than Model B.



**Supplemental figure S2.** Model AUC (500 Days) when Applied to All Cases from the Test Partition. This figure is analogous to **Figure 3**, but shows AUC values at 500 days. See the Figure 3 legend for additional detail.



**Supplemental figure S3.** Model AUC (1500 Days) when Applied to All Cases from the Test Partition. This figure is analogous to **Figure 3**, but shows AUC values at 1500 days. See the Figure 3 legend for additional detail.



**Supplemental figure S4.** Colorectal Cancer Meta-model Performance Using Various Weighting Functions. The meta-model is based on a weighted average of predictions from underlying models. Weights are assigned according to the equation  $w = (x - 0.5)^n$ , where w is the weight assigned to the model prediction, x represents the median cross validation AUC and n represents an exponent (0.5 is intended to represent the AUC of the null model). While for most of our analysis, we used n=2, we tested the impact of various values of n as shown above. Performance represents meta-model AUC when applied to all patients within the test partition. The actual value of n had only a modest impact.

**Supplemental table S1.** Detailed Model Performance Results

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Breast Cancer	Cox	A	CompleteCases
Breast Cancer	Cox	B	CompleteCases
Breast Cancer	Cox	C	CompleteCases
Breast Cancer	Cox	D	CompleteCases
Breast Cancer	Cox	E	CompleteCases
Breast Cancer	MetaModel	A	MetaModel
Breast Cancer	MetaModel	all	MetaModel
Breast Cancer	MetaModel	B	MetaModel
Breast Cancer	MetaModel	C	MetaModel
Breast Cancer	MetaModel	D	MetaModel
Breast Cancer	MetaModel	E	MetaModel
Breast Cancer	RF	A	CompleteCases
Breast Cancer	RF	A	CompleteCases
Breast Cancer	RF	A	CompleteCases
Breast Cancer	RF	A	missForest
Breast Cancer	RF	A	missForest
Breast Cancer	RF	A	missForest
Breast Cancer	RF	A	NativeImputation
Breast Cancer	RF	A	NativeImputation
Breast Cancer	RF	A	NativeImputation
Breast Cancer	RF	B	CompleteCases
Breast Cancer	RF	B	CompleteCases
Breast Cancer	RF	B	CompleteCases
Breast Cancer	RF	B	missForest
Breast Cancer	RF	B	missForest
Breast Cancer	RF	B	NativeImputation
Breast Cancer	RF	B	NativeImputation
Breast Cancer	RF	B	NativeImputation
Breast Cancer	RF	C	CompleteCases
Breast Cancer	RF	C	CompleteCases
Breast Cancer	RF	C	CompleteCases
Breast Cancer	RF	C	missForest
Breast Cancer	RF	C	missForest
Breast Cancer	RF	C	NativeImputation
Breast Cancer	RF	C	NativeImputation
Breast Cancer	RF	C	NativeImputation
Breast Cancer	RF	D	CompleteCases
Breast Cancer	RF	D	CompleteCases
Breast Cancer	RF	D	CompleteCases

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
CompleteCases	0.52	0.56	0.53	0.54	0.54	0.55
CompleteCases	0.59	0.59	0.58	0.59	0.58	0.59
CompleteCases	0.60	0.61	0.59	0.61	0.60	0.61
CompleteCases	0.72	0.71	0.70	0.69	0.67	0.68
CompleteCases	0.79	0.78	0.75	0.76	0.73	0.74
MetaModel	0.74	0.87	0.74	0.85	0.72	0.86
MetaModel	0.74	NA	0.74	NA	0.72	NA
MetaModel	0.74	0.87	0.74	0.85	0.72	0.86
MetaModel	0.74	0.87	0.74	0.85	0.72	0.86
MetaModel	0.75	0.89	0.75	0.86	0.72	0.87
MetaModel	0.78	0.93	0.77	0.91	0.75	0.92
CompleteCases	0.53	0.60	0.55	0.59	0.55	0.59
missForest	0.53	NA	0.55	NA	0.55	NA
NativeImputation	0.53	NA	0.55	NA	0.55	NA
CompleteCases	0.54	NA	0.55	NA	0.55	NA
missForest	0.54	0.60	0.55	0.59	0.55	0.59
NativeImputation	0.54	NA	0.55	NA	0.55	NA
CompleteCases	0.54	NA	0.55	NA	0.55	NA
missForest	0.54	NA	0.55	NA	0.55	NA
NativeImputation	0.54	0.60	0.55	0.59	0.55	0.59
CompleteCases	0.59	0.59	0.58	0.59	0.59	0.59
missForest	0.59	NA	0.58	NA	0.58	NA
NativeImputation	0.58	NA	0.58	NA	0.59	NA
CompleteCases	0.59	NA	0.58	NA	0.58	NA
missForest	0.59	0.59	0.58	0.59	0.58	0.59
NativeImputation	0.60	NA	0.58	NA	0.58	NA
CompleteCases	0.59	NA	0.57	NA	0.59	NA
missForest	0.59	NA	0.58	NA	0.58	NA
NativeImputation	0.60	0.59	0.58	0.59	0.59	0.59
CompleteCases	0.55	0.68	0.56	0.67	0.57	0.67
missForest	0.55	NA	0.56	NA	0.57	NA
NativeImputation	0.55	NA	0.56	NA	0.57	NA
CompleteCases	0.55	NA	0.56	NA	0.57	NA
missForest	0.55	0.68	0.56	0.67	0.57	0.67
NativeImputation	0.55	NA	0.56	NA	0.57	NA
CompleteCases	0.54	NA	0.56	NA	0.57	NA
missForest	0.55	NA	0.56	NA	0.57	NA
NativeImputation	0.55	0.68	0.56	0.67	0.57	0.67
CompleteCases	0.68	0.80	0.69	0.78	0.66	0.78
missForest	0.68	NA	0.69	NA	0.66	NA
NativeImputation	0.67	NA	0.69	NA	0.66	NA

**Supplemental table S1.** (continued)

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Breast Cancer	RF	D	missForest
Breast Cancer	RF	D	missForest
Breast Cancer	RF	D	missForest
Breast Cancer	RF	D	NativeImputation
Breast Cancer	RF	D	NativeImputation
Breast Cancer	RF	D	NativeImputation
Breast Cancer	RF	E	CompleteCases
Breast Cancer	RF	E	CompleteCases
Breast Cancer	RF	E	CompleteCases
Breast Cancer	RF	E	missForest
Breast Cancer	RF	E	missForest
Breast Cancer	RF	E	missForest
Breast Cancer	RF	E	NativeImputation
Breast Cancer	RF	E	NativeImputation
Breast Cancer	RF	E	NativeImputation
Colorectal Cancer	Cox	A	CompleteCases
Colorectal Cancer	Cox	B	CompleteCases
Colorectal Cancer	Cox	C	CompleteCases
Colorectal Cancer	Cox	D	CompleteCases
Colorectal Cancer	Cox	E	CompleteCases
Colorectal Cancer	Cox	F	CompleteCases
Colorectal Cancer	Cox	G	CompleteCases
Colorectal Cancer	MetaModel	A	MetaModel
Colorectal Cancer	MetaModel	all	MetaModel
Colorectal Cancer	MetaModel	B	MetaModel
Colorectal Cancer	MetaModel	C	MetaModel
Colorectal Cancer	MetaModel	D	MetaModel
Colorectal Cancer	MetaModel	E	MetaModel
Colorectal Cancer	MetaModel	F	MetaModel
Colorectal Cancer	MetaModel	G	MetaModel
Colorectal Cancer	RF	A	CompleteCases
Colorectal Cancer	RF	A	CompleteCases
Colorectal Cancer	RF	A	CompleteCases
Colorectal Cancer	RF	A	missForest
Colorectal Cancer	RF	A	missForest
Colorectal Cancer	RF	A	missForest
Colorectal Cancer	RF	A	NativeImputation
Colorectal Cancer	RF	A	NativeImputation
Colorectal Cancer	RF	A	NativeImputation
Colorectal Cancer	RF	B	CompleteCases
Colorectal Cancer	RF	B	CompleteCases

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
CompleteCases	0.67	NA	0.68	NA	0.66	NA
missForest	0.67	0.79	0.69	0.77	0.66	0.78
NativeImputation	0.67	NA	0.68	NA	0.65	NA
CompleteCases	0.67	NA	0.68	NA	0.65	NA
missForest	0.68	NA	0.69	NA	0.66	NA
NativeImputation	0.67	0.80	0.68	0.78	0.65	0.79
CompleteCases	0.77	0.94	0.77	0.92	0.76	0.93
missForest	0.77	NA	0.75	NA	0.73	NA
NativeImputation	0.75	NA	0.73	NA	0.70	NA
CompleteCases	0.78	NA	0.77	NA	0.76	NA
missForest	0.78	0.92	0.76	0.91	0.72	0.92
NativeImputation	0.77	NA	0.75	NA	0.71	NA
CompleteCases	0.78	NA	0.77	NA	0.76	NA
missForest	0.78	NA	0.76	NA	0.73	NA
NativeImputation	0.76	0.95	0.74	0.94	0.71	0.95
CompleteCases	0.64	0.61	0.59	0.59	0.59	0.58
CompleteCases	0.57	0.58	0.58	0.60	0.58	0.61
CompleteCases	0.65	0.63	0.63	0.63	0.63	0.64
CompleteCases	0.75	0.75	0.73	0.73	0.73	0.74
CompleteCases	0.74	0.76	0.72	0.74	0.73	0.74
CompleteCases	0.70	0.71	0.68	0.69	0.67	0.70
CompleteCases	0.71	0.74	0.65	0.72	0.59	0.73
MetaModel	0.76	0.88	0.71	0.88	0.72	0.89
MetaModel	0.76	NA	0.71	NA	0.72	NA
MetaModel	0.76	0.88	0.71	0.88	0.72	0.89
MetaModel	0.76	0.88	0.71	0.88	0.72	0.89
MetaModel	0.79	0.94	0.77	0.94	0.78	0.96
MetaModel	0.78	0.95	0.76	0.94	0.78	0.96
MetaModel	0.76	0.93	0.76	0.93	0.77	0.96
MetaModel	0.79	0.95	0.72	0.93	0.71	0.95
CompleteCases	0.62	0.64	0.57	0.63	0.56	0.63
missForest	0.62	NA	0.57	NA	0.56	NA
NativeImputation	0.62	NA	0.56	NA	0.56	NA
CompleteCases	0.62	NA	0.57	NA	0.56	NA
missForest	0.62	0.64	0.57	0.63	0.56	0.63
NativeImputation	0.62	NA	0.57	NA	0.56	NA
CompleteCases	0.61	NA	0.56	NA	0.56	NA
missForest	0.61	NA	0.56	NA	0.56	NA
NativeImputation	0.61	0.64	0.56	0.63	0.56	0.63
CompleteCases	0.58	0.60	0.57	0.61	0.58	0.62
missForest	0.58	NA	0.57	NA	0.58	NA

**Supplemental table S1.** (continued)

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Colorectal Cancer	RF	B	CompleteCases
Colorectal Cancer	RF	B	missForest
Colorectal Cancer	RF	B	missForest
Colorectal Cancer	RF	B	missForest
Colorectal Cancer	RF	B	NativeImputation
Colorectal Cancer	RF	B	NativeImputation
Colorectal Cancer	RF	B	NativeImputation
Colorectal Cancer	RF	C	CompleteCases
Colorectal Cancer	RF	C	CompleteCases
Colorectal Cancer	RF	C	missForest
Colorectal Cancer	RF	C	missForest
Colorectal Cancer	RF	C	NativeImputation
Colorectal Cancer	RF	C	NativeImputation
Colorectal Cancer	RF	C	NativeImputation
Colorectal Cancer	RF	D	CompleteCases
Colorectal Cancer	RF	D	CompleteCases
Colorectal Cancer	RF	D	CompleteCases
Colorectal Cancer	RF	D	missForest
Colorectal Cancer	RF	D	missForest
Colorectal Cancer	RF	D	missForest
Colorectal Cancer	RF	D	NativeImputation
Colorectal Cancer	RF	D	NativeImputation
Colorectal Cancer	RF	E	CompleteCases
Colorectal Cancer	RF	E	CompleteCases
Colorectal Cancer	RF	E	missForest
Colorectal Cancer	RF	E	missForest
Colorectal Cancer	RF	E	missForest
Colorectal Cancer	RF	E	NativeImputation
Colorectal Cancer	RF	E	NativeImputation
Colorectal Cancer	RF	E	NativeImputation
Colorectal Cancer	RF	F	CompleteCases
Colorectal Cancer	RF	F	CompleteCases
Colorectal Cancer	RF	F	CompleteCases
Colorectal Cancer	RF	F	missForest
Colorectal Cancer	RF	F	missForest
Colorectal Cancer	RF	F	missForest
Colorectal Cancer	RF	F	NativeImputation

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
NativeImputation	0.58	NA	0.57	NA	0.58	NA
CompleteCases	0.58	NA	0.57	NA	0.58	NA
missForest	0.58	0.60	0.57	0.61	0.58	0.62
NativeImputation	0.58	NA	0.57	NA	0.58	NA
CompleteCases	0.57	NA	0.56	NA	0.58	NA
missForest	0.57	NA	0.56	NA	0.58	NA
NativeImputation	0.58	0.60	0.57	0.61	0.58	0.62
CompleteCases	0.63	0.73	0.60	0.74	0.61	0.76
missForest	0.63	NA	0.60	NA	0.61	NA
NativeImputation	0.63	NA	0.60	NA	0.61	NA
CompleteCases	0.63	NA	0.61	NA	0.61	NA
missForest	0.63	0.74	0.61	0.74	0.61	0.76
NativeImputation	0.63	NA	0.61	NA	0.61	NA
CompleteCases	0.64	NA	0.61	NA	0.62	NA
missForest	0.64	NA	0.61	NA	0.62	NA
NativeImputation	0.64	0.73	0.61	0.74	0.62	0.76
CompleteCases	0.76	0.93	0.72	0.93	0.75	0.95
missForest	0.75	NA	0.69	NA	0.69	NA
NativeImputation	0.72	NA	0.67	NA	0.70	NA
CompleteCases	0.76	NA	0.72	NA	0.75	NA
missForest	0.75	0.92	0.70	0.92	0.70	0.94
NativeImputation	0.73	NA	0.68	NA	0.70	NA
CompleteCases	0.77	NA	0.72	NA	0.74	NA
missForest	0.75	NA	0.69	NA	0.70	NA
NativeImputation	0.73	0.95	0.68	0.95	0.70	0.97
CompleteCases	0.77	0.93	0.75	0.93	0.76	0.95
missForest	0.75	NA	0.70	NA	0.72	NA
NativeImputation	0.74	NA	0.70	NA	0.73	NA
CompleteCases	0.76	NA	0.73	NA	0.76	NA
missForest	0.76	0.92	0.71	0.92	0.72	0.94
NativeImputation	0.74	NA	0.70	NA	0.73	NA
CompleteCases	0.77	NA	0.75	NA	0.77	NA
missForest	0.76	NA	0.70	NA	0.72	NA
NativeImputation	0.74	0.96	0.70	0.96	0.73	0.97
CompleteCases	0.75	0.91	0.73	0.91	0.74	0.94
missForest	0.74	NA	0.70	NA	0.71	NA
NativeImputation	0.73	NA	0.70	NA	0.72	NA
CompleteCases	0.75	NA	0.73	NA	0.74	NA
missForest	0.76	0.91	0.72	0.91	0.73	0.93
NativeImputation	0.74	NA	0.71	NA	0.73	NA
CompleteCases	0.75	NA	0.74	NA	0.74	NA

**Supplemental table S1.** (continued)

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Colorectal Cancer	RF	F	NativeImputation
Colorectal Cancer	RF	F	NativeImputation
Colorectal Cancer	RF	G	CompleteCases
Colorectal Cancer	RF	G	CompleteCases
Colorectal Cancer	RF	G	CompleteCases
Colorectal Cancer	RF	G	missForest
Colorectal Cancer	RF	G	missForest
Colorectal Cancer	RF	G	missForest
Colorectal Cancer	RF	G	NativeImputation
Colorectal Cancer	RF	G	NativeImputation
Colorectal Cancer	RF	G	NativeImputation
Lung Cancer	Cox	A	CompleteCases
Lung Cancer	Cox	B	CompleteCases
Lung Cancer	Cox	C	CompleteCases
Lung Cancer	Cox	D	CompleteCases
Lung Cancer	Cox	E	CompleteCases
Lung Cancer	Cox	F	CompleteCases
Lung Cancer	Cox	G	CompleteCases
Lung Cancer	Cox	H	CompleteCases
Lung Cancer	Cox	rn	CompleteCases
Lung Cancer	MetaModel	A	MetaModel
Lung Cancer	MetaModel	all	MetaModel
Lung Cancer	MetaModel	B	MetaModel
Lung Cancer	MetaModel	C	MetaModel
Lung Cancer	MetaModel	D	MetaModel
Lung Cancer	MetaModel	E	MetaModel
Lung Cancer	MetaModel	F	MetaModel
Lung Cancer	MetaModel	G	MetaModel
Lung Cancer	MetaModel	H	MetaModel
Lung Cancer	RF	A	CompleteCases
Lung Cancer	RF	A	CompleteCases
Lung Cancer	RF	A	CompleteCases
Lung Cancer	RF	A	missForest
Lung Cancer	RF	A	missForest
Lung Cancer	RF	A	missForest
Lung Cancer	RF	A	NativeImputation
Lung Cancer	RF	A	NativeImputation
Lung Cancer	RF	A	NativeImputation
Lung Cancer	RF	B	CompleteCases
Lung Cancer	RF	B	CompleteCases
Lung Cancer	RF	B	CompleteCases

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
missForest	0.75	NA	0.70	NA	0.72	NA
NativeImputation	0.74	0.92	0.71	0.93	0.73	0.96
CompleteCases	0.76	0.93	0.71	0.91	0.68	0.92
missForest	0.73	NA	0.69	NA	0.70	NA
NativeImputation	0.74	NA	0.70	NA	0.71	NA
CompleteCases	0.78	NA	0.71	NA	0.71	NA
missForest	0.76	0.91	0.71	0.91	0.72	0.93
NativeImputation	0.75	NA	0.70	NA	0.72	NA
CompleteCases	0.77	NA	0.71	NA	0.69	NA
missForest	0.75	NA	0.70	NA	0.72	NA
NativeImputation	0.74	0.95	0.70	0.94	0.72	0.96
CompleteCases	0.58	0.58	0.60	0.58	0.60	0.59
CompleteCases	0.65	0.63	0.65	0.65	0.65	0.65
CompleteCases	0.66	0.65	0.68	0.67	0.68	0.68
CompleteCases	0.70	0.71	0.72	0.71	0.70	0.72
CompleteCases	0.57	0.58	0.56	0.58	0.54	0.58
CompleteCases	0.68	0.68	0.69	0.69	0.68	0.69
CompleteCases	0.73	0.73	0.75	0.72	0.74	0.73
CompleteCases	0.64	0.84	0.67	0.81	0.62	0.83
CompleteCases	0.51	0.50	0.50	0.52	0.49	0.52
MetaModel	0.72	0.88	0.73	0.89	0.72	0.91
MetaModel	0.72	NA	0.73	NA	0.72	NA
MetaModel	0.72	0.88	0.73	0.89	0.72	0.91
MetaModel	0.72	0.88	0.73	0.89	0.72	0.91
MetaModel	0.74	0.92	0.74	0.93	0.72	0.95
MetaModel	0.71	0.88	0.74	0.88	0.72	0.90
MetaModel	0.71	0.88	0.74	0.88	0.72	0.90
MetaModel	0.74	0.91	0.75	0.93	0.72	0.95
MetaModel	0.78	0.96	0.88	0.96	0.77	0.95
CompleteCases	0.54	0.61	0.59	0.61	0.59	0.63
missForest	0.54	NA	0.59	NA	0.59	NA
NativeImputation	0.54	NA	0.59	NA	0.59	NA
CompleteCases	0.54	NA	0.59	NA	0.59	NA
missForest	0.54	0.61	0.59	0.61	0.59	0.63
NativeImputation	0.54	NA	0.59	NA	0.59	NA
CompleteCases	0.54	NA	0.60	NA	0.58	NA
missForest	0.54	NA	0.59	NA	0.58	NA
NativeImputation	0.54	0.61	0.60	0.61	0.58	0.63
CompleteCases	0.66	0.64	0.65	0.66	0.66	0.66
missForest	0.65	NA	0.64	NA	0.65	NA
NativeImputation	0.65	NA	0.65	NA	0.65	NA

**Supplemental table S1.** (continued)

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Lung Cancer	RF	B	missForest
Lung Cancer	RF	B	missForest
Lung Cancer	RF	B	missForest
Lung Cancer	RF	B	NativeImputation
Lung Cancer	RF	B	NativeImputation
Lung Cancer	RF	B	NativeImputation
Lung Cancer	RF	C	CompleteCases
Lung Cancer	RF	C	CompleteCases
Lung Cancer	RF	C	CompleteCases
Lung Cancer	RF	C	missForest
Lung Cancer	RF	C	missForest
Lung Cancer	RF	C	missForest
Lung Cancer	RF	C	NativeImputation
Lung Cancer	RF	C	NativeImputation
Lung Cancer	RF	C	NativeImputation
Lung Cancer	RF	D	CompleteCases
Lung Cancer	RF	D	CompleteCases
Lung Cancer	RF	D	CompleteCases
Lung Cancer	RF	D	missForest
Lung Cancer	RF	D	missForest
Lung Cancer	RF	D	NativeImputation
Lung Cancer	RF	D	NativeImputation
Lung Cancer	RF	E	CompleteCases
Lung Cancer	RF	E	CompleteCases
Lung Cancer	RF	E	CompleteCases
Lung Cancer	RF	E	missForest
Lung Cancer	RF	E	missForest
Lung Cancer	RF	E	missForest
Lung Cancer	RF	E	NativeImputation
Lung Cancer	RF	E	NativeImputation
Lung Cancer	RF	E	NativeImputation
Lung Cancer	RF	F	CompleteCases
Lung Cancer	RF	F	CompleteCases
Lung Cancer	RF	F	CompleteCases
Lung Cancer	RF	F	missForest
Lung Cancer	RF	F	missForest
Lung Cancer	RF	F	missForest
Lung Cancer	RF	F	NativeImputation
Lung Cancer	RF	F	NativeImputation

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
CompleteCases	0.65	NA	0.65	NA	0.65	NA
missForest	0.65	0.64	0.65	0.66	0.66	0.66
NativeImputation	0.65	NA	0.65	NA	0.65	NA
CompleteCases	0.65	NA	0.65	NA	0.65	NA
missForest	0.65	NA	0.65	NA	0.65	NA
NativeImputation	0.65	0.64	0.65	0.66	0.66	0.66
CompleteCases	0.64	0.73	0.67	0.76	0.65	0.78
missForest	0.64	NA	0.67	NA	0.65	NA
NativeImputation	0.64	NA	0.67	NA	0.65	NA
CompleteCases	0.64	NA	0.67	NA	0.64	NA
missForest	0.64	0.73	0.67	0.76	0.64	0.78
NativeImputation	0.64	NA	0.67	NA	0.64	NA
CompleteCases	0.64	NA	0.67	NA	0.65	NA
missForest	0.64	NA	0.67	NA	0.65	NA
NativeImputation	0.64	0.73	0.67	0.76	0.65	0.78
CompleteCases	0.71	0.91	0.72	0.94	0.70	0.96
missForest	0.71	NA	0.72	NA	0.71	NA
NativeImputation	0.67	NA	0.69	NA	0.67	NA
CompleteCases	0.71	NA	0.72	NA	0.70	NA
missForest	0.72	0.90	0.72	0.92	0.72	0.94
NativeImputation	0.70	NA	0.70	NA	0.69	NA
CompleteCases	0.71	NA	0.73	NA	0.71	NA
missForest	0.72	NA	0.73	NA	0.71	NA
NativeImputation	0.70	0.94	0.71	0.95	0.69	0.97
CompleteCases	0.58	0.59	0.56	0.58	0.57	0.59
missForest	0.56	NA	0.57	NA	0.55	NA
NativeImputation	0.51	NA	0.52	NA	0.53	NA
CompleteCases	0.58	NA	0.58	NA	0.57	NA
missForest	0.56	0.58	0.57	0.57	0.55	0.57
NativeImputation	0.52	NA	0.54	NA	0.54	NA
CompleteCases	0.58	NA	0.57	NA	0.57	NA
missForest	0.57	NA	0.55	NA	0.56	NA
NativeImputation	0.53	0.59	0.54	0.58	0.53	0.59
CompleteCases	0.69	0.77	0.69	0.77	0.67	0.79
missForest	0.67	NA	0.66	NA	0.63	NA
NativeImputation	0.66	NA	0.66	NA	0.63	NA
CompleteCases	0.66	NA	0.69	NA	0.66	NA
missForest	0.66	0.76	0.68	0.78	0.66	0.81
NativeImputation	0.65	NA	0.68	NA	0.66	NA
CompleteCases	0.67	NA	0.69	NA	0.66	NA
missForest	0.66	NA	0.68	NA	0.66	NA

**Supplemental table S1.** (continued)

<b>Dataset</b>	<b>Model Type</b>	<b>Predictor Set</b>	<b>Missing Training Data Addressed</b>
Lung Cancer	RF	F	NativeImputation
Lung Cancer	RF	G	CompleteCases
Lung Cancer	RF	G	CompleteCases
Lung Cancer	RF	G	CompleteCases
Lung Cancer	RF	G	missForest
Lung Cancer	RF	G	missForest
Lung Cancer	RF	G	missForest
Lung Cancer	RF	G	NativeImputation
Lung Cancer	RF	G	NativeImputation
Lung Cancer	RF	G	NativeImputation
Lung Cancer	RF	H	CompleteCases
Lung Cancer	RF	H	CompleteCases
Lung Cancer	RF	H	CompleteCases
Lung Cancer	RF	H	missForest
Lung Cancer	RF	H	missForest
Lung Cancer	RF	H	missForest
Lung Cancer	RF	H	NativeImputation
Lung Cancer	RF	H	NativeImputation
Lung Cancer	RF	H	NativeImputation

## SUPPLEMENTAL METHODS

This section provides additional methodologic detail to supplement the methods in the main text.

### Patient Cohort Definitions

The metastatic CRC and breast cancer cohorts included all patients treated in the Flatiron Health network for which data was available who had a first metastasis diagnosed during the 2013 or 2014 calendar years. Likewise, the NSCLC cohort included all available patients with lung cancer in the corresponding DataMart who had an advanced diagnosis (stage III(B) or higher) in 2013. Cases and data-elements were included if available in a data extract released June, 2020.

### Time Point Normalization

We normalized the time of each predictor or outcome data element to the date of first metastasis for CRC and breast cancer patients or the date of advanced diagnosis for lung cancer patients.

Missing Testing Data Addressed	AUC					
	500 Days		1000 Days		1500 Days	
	Testing	Training	Testing	Training	Testing	Training
NativeImputation	0.66	0.75	0.68	0.77	0.66	0.80
CompleteCases	0.70	0.92	0.74	0.94	0.72	0.96
missForest	0.70	NA	0.72	NA	0.71	NA
NativeImputation	0.67	NA	0.69	NA	0.67	NA
CompleteCases	0.71	NA	0.73	NA	0.70	NA
missForest	0.72	0.90	0.72	0.92	0.71	0.94
NativeImputation	0.70	NA	0.71	NA	0.69	NA
CompleteCases	0.72	NA	0.76	NA	0.74	NA
missForest	0.73	NA	0.74	NA	0.74	NA
NativeImputation	0.70	0.94	0.72	0.95	0.70	0.97
CompleteCases	0.76	0.96	0.84	0.97	0.72	0.95
missForest	0.65	NA	0.68	NA	0.67	NA
NativeImputation	0.64	NA	0.67	NA	0.65	NA
CompleteCases	0.78	NA	0.81	NA	0.68	NA
missForest	0.73	0.91	0.72	0.93	0.72	0.95
NativeImputation	0.71	NA	0.70	NA	0.69	NA
CompleteCases	0.77	NA	0.86	NA	0.74	NA
missForest	0.73	NA	0.74	NA	0.74	NA
NativeImputation	0.71	0.97	0.72	0.97	0.70	0.98

## Defining Survival (outcome variable)

We defined "survival" as follows:

1. For patients with a date of death listed in the database, we defined survival as the number of days between the patient's advanced diagnosis date and the date of death.
2. For patients without a listed date of death, we treated patients as alive up until the time of the last recorded visit (which we captured in terms of number of days since advanced diagnosis) and "censored" patients from analysis for all time points thereafter. Rare patients had a date of death that was less than or equal to the date of advanced diagnosis; these patients were excluded due to the data inconsistency.

## Predictor Variable Pre-processing

Available results for selected quantitative laboratory tests were extracted for each patient in each cohort for specimens collected around the time (within +/- 90 days) of advanced diagnosis. Test results were transformed into absolute median deviations (AMD) minimum and maximum AMD values for each patient during the +/- 90 day time window were

included as predictors in our models. We chose to use absolute median deviation instead of the values themselves to reflect the notion that for many tests, high or low values can be diagnostically informative.

## Other Predictors

We captured the patient's age (at date of advanced diagnosis) based on the patient's birth year (ages were calculated assuming a January 1<sup>st</sup> birthday), the patient's gender and the initial tumor stage. For the CRC cohort, we additionally captured the KRAS and BRAF status and the tumor site (e.g. colon). For the breast cancer cohort, we additionally captured the patient's ER, PR and HER2 status and for the lung cancer cohort we captured the ALK and EGFR mutation status. For the molecular biomarkers (ER, PR, HER2, KRAS, BRAF, EGFR, ALK), we considered only the result for each marker on each patient nearest in time (based on specimen collect date; ties broken randomly) to the advanced diagnosis date. We then classified the molecular biomarker results as "Positive", "Negative" or "Other" with "positive" indicating that the corresponding marker was expressed or a mutation was present. A "negative" classification indicated that the marker was absent or the mutation was not observed and "other" indicated an equivocal result, a test that was not able to be performed or other non-definitive result. We developed a custom grouper to translate raw molecular biomarker results into the classifications described above.

## Survival Models

Technical details of the linear and random forest-based survival model development algorithms have been described previously (22 34). Briefly, the linear models assume that a patient's hazard (instantaneous risk of death) at any point in time (in this particular cases, days since advanced diagnosis) is in a constant (across all time points) proportion to the "baseline" hazard curve, where the baseline hazard curve represents that hazard at each time point experienced by the average patient in the cohort. The patient-specific hazard proportion (by which we multiplied the baseline hazard to arrive at the patient-specific hazard) was calculated as a linear combination of predictors with the specific coefficients fit based on maximum likelihood in accordance with the standard Cox regression approach. Random survival forests predict a patient-specific hazard at each time point that is not necessarily in proportion to baseline hazard (e.g. a given factor could lead to large hazard later in the course). Each model includes a "forest" of individual decision trees where each tree is built from a randomly selected subset of cases (from the training data) and predictor features. The approach described above generated predicted cumulative hazard functions for patients; we transformed these into predicted survival using the well-established formula:  $S=\exp(-H)$ , where  $S$  is the probability of survival and  $H$  is the cumulative hazard.

## Calculation of AUC Statistics

We derived AUC values using the method of Hagerty et al. (28) implemented in the R package survivalROC. This method utilizes both censored and uncensored cases. It derives estimates of the model's sensitivity and specificity at a given time point and model predicted survival threshold (where patients with a predicted survival less than the threshold would be classified as deceased) by comparing model predicted survivals to actual patient survivals estimated using either the Kaplan-Meier or nearest neighbor method. For our calculations, we used the Kaplan-Meier method. We considered cut values for calculating sensitivity and specificity in one-percentile increments across the range of predicted survival probabilities. (We used 5 percentile increments for meta-model cross validation to improve speed).

## Meta-model weighting

As noted in the methods, we assigned a weight to each individual model based on cross validation performance. To do this, we performed five-fold cross validation of each individual model capturing the median cross validation AUC of the model at 500, 1000 and 1500 days. (The overall cross validation AUC for each underlying model was a median across 15 values; 5 cross-validation runs times 3 time points). We then transformed these median AUC values into model weights using one of several weighting functions having the form  $w = (x - 0.5)^n$ , where w is the weight assigned to the model prediction, x represents the median cross validation AUC and n represents an exponent (0.5 is intended to represent the AUC of the null model). A value of  $n=0$  is equivalent to an unweighted average (all underlying models are weighted equally) while the larger the value of n, the more emphasis the meta-model places on the better performing underlying models. We used a single weight for each model across all time points. We explored use of values of n from 0 to 3.

When applying weights to average individual model predictors for a patient, we normalized weights so that the weights for all models used for a given patient sum to one. In theory a model producing an AUC <0.5 would produce a negative weight; while not generally expected to be an issue in practice, negatively weighting outputs from these models would in theory be expected to improve prediction accuracy (since an AUC <0.5 would indicate that unfavorable model predicted prognosis actually indicates better actual prognosis).

## Calibration

To evaluate whether the models produced output that was well calibrated, we considered whether the predicted probabilities of survival matched actual survival overall groups of patients and time windows. To do this, we considered time in 250 day time windows from 0-1500 days (6 windows total). For each patient, we estimated a predicted likelihood of death during the time window based on the model's predicted survival for the patient at the start and end of the time interval, using the equation:

$$P_d = 1 - \frac{S_{end}}{S_{start}},$$

where  $P_d$  is the patient's predicted risk of death during the time window conditioned on surviving to the start of the interval, and  $S_{end}$  is the predicted survival at the end of the time window;  $S_{start}$  normalizes likelihood that the patient did not survive to the start of the time interval. Patients were excluded from a time window if they died or were censored prior to the start of the time window; patients censored during the time window were included, but  $S_{end}$  was based on the predicted survival at the time of censoring instead of the end of the window.

After calculating predicted likelihood of death for each patient during the time window, we grouped patients into quartiles from lowest to highest predicted risk of death during the time window. We summed the predicted risk across patients in each time-window-risk quartile and counted the number of deaths to calculate both predicted and actual deaths for the patient group during the time window (See Figure 4, top). We compared the predicted deaths to actual deaths for each patient-risk-quartile-time-window combination.

To summarize each model's overall calibration, we calculated the slope of Poisson regression model (fit using the `glm` function in R), predicting actual deaths as a function of the predicted deaths. More specifically, we estimated the term  $\beta$  in the equation:

$$\log[E(Y|X)] = \beta X,$$

where Y is the actual deaths for the patients in a risk quartile during a time window, X is the log of predicted deaths (calculated as above) for the corresponding patients and time window,  $\beta$  is the slope being fit and E represents the expectation. Slopes (values of  $\beta$ ) near one indicate close alignment of predicted and actual deaths and suggest a well-calibrated model. This procedure was adapted based on previously described methods including those in Rahman et al. (35)

### **missForest Imputation**

As discussed in the methods, we performed imputation of missing predictor values using missForest for comparison to the meta-models. When using missForest, we performed imputation separately for each dataset and partition (i.e. six imputations, one for the training and one for the testing partition for each of the three datasets corresponding to the three cancer types). We imputed values for and using available results for any predictor that was included in any predictor set for the cohort. We set the maximum imputation iterations parameter at 10 and the number of trees at 100.

## General Computation

Data were extracted using SQL queries from a Teradata database. Except as otherwise specified, all key computational steps were performed in R (36) and all key plots were generated using the R ggplots2 package.

## Supplemental Table S2: Additional Calibration Data and Associated Methods

As described in the main manuscript, to test model calibration, we calculated each test patient's predicted mortality over 250 day time windows. We further grouped patients into risk quartiles (4= highest risk of dying; 1= lowest risk) and calculated the aggregate predicted mortality for each patient- group, time-window combination. We then compared predicted mortality to actual mortality. To summarize each model's calibration, we calculated the slope of a Poisson regression line fitting actual- mortality as a function of (log-transformed) predicted mortality with an intercept through the origin.

In addition to the analysis provided in the main text and in Figure 4, here we provide additional analysis to assess the impact of the predicted risk quartile on calibration. For this additional analysis, we include an additional predictor in the Poisson regression model indicating the patient's predicted risk quartile (1,2,3 or 4 where 1= lowest risk, 4= highest risk; we subtract 2.5 from these scores within the regression model to center at zero, since we want to force the regression line through the origin).

In the table below, the columns labeled "Unadjusted Calibration" show coefficients for log-transformed predicted mortality in regression models that do not explicitly include the risk quartile (although data points are nonetheless grouped into risk quartiles). The values in this column are methodologically identical to those provide in Figure 4.\* The columns labeled "Adjusted Calibration" provide the coefficients for log-transformed predicted mortality in regression models that do explicitly include the risk quartile.

Coefficient values near 1 indicate optimal model calibration. As shown, explicitly including the risk quartile in the model generally seems to have limited impact on apparent calibration, suggesting that the models do not tend to systematically over or under estimate mortality in high or low risk patients.

\* Please note that data for this Supplemental Table S2 was generated by revising and rerunning the analysis script weeks following the analysis used to generate all other data in this manuscript. While not material to overall manuscript findings and conclusions, there were minor differences in the results from this rerun analysis as compared to the initial run. These differences are suspected to be due to updates in the R analysis packages used and/or by calls to the random number generator (e.g. for splitting into training and testing or for model fitting). Nonetheless, the coefficient values here will not exactly match those in Figure 4. Thus, this table should only be used to compare adjusted to unadjusted calibration values and the data shown in this table should not be directly compared to other data in this manuscript.

Training Data	Model	Testing Data Imputation: missForest		Testing Data Imputation: NativeImputation	
		Unadjusted Calibration	Adjusted Calibration	Unadjusted Calibration	Adjusted Calibration
CompleteCases	Br_A	1.00	1.02	1.00	1.02
	Br_B	1.00	1.00	1.00	0.99
	Br_C	0.99	1.05	0.99	1.05
	Br_D	0.99	1.05	0.99	1.04
	Br_E	0.99	0.98	0.94	0.92
	CRC_A	1.01	1.03	1.01	1.03
	CRC_B	1.01	1.01	1.01	1.01
	CRC_C	1.01	1.04	1.01	1.04
	CRC_D	1.00	1.00	0.96	0.95
	CRC_E	0.98	0.97	0.96	0.96
	CRC_F	1.00	0.99	0.99	0.98
	CRC_G	1.07	1.07	1.05	1.04
	ICA_A	1.00	1.01	1.00	1.01
	ICA_B	1.01	1.01	1.01	1.01
	ICA_C	1.00	1.03	1.00	1.03
	ICA_D	0.98	0.98	0.97	0.97
	ICA_E	1.03	1.04	1.04	1.04
	ICA_F	1.04	1.06	1.04	1.06
	ICA_G	1.01	1.01	0.99	0.99
	ICA_H	1.05	1.07	1.03	1.04
missForest	Br_A	1.00	1.02	1.00	1.02
	Br_B	0.99	0.99	0.99	0.99
	Br_C	0.99	1.05	0.99	1.05
	Br_D	0.98	1.04	0.98	1.04
	Br_E	0.97	0.95	0.95	0.92
	CRC_A	1.01	1.03	1.01	1.03
	CRC_B	1.01	1.02	1.01	1.02
	CRC_C	1.01	1.04	1.01	1.04
	CRC_D	1.00	0.99	0.97	0.96
	CRC_E	1.00	0.98	0.98	0.95
	CRC_F	1.00	0.98	0.98	0.96
	CRC_G	1.00	0.98	0.98	0.96
	ICA_A	1.00	1.01	1.00	1.01
	ICA_B	1.00	1.01	1.00	1.01
	ICA_C	1.00	1.03	1.00	1.03
	ICA_D	0.99	0.99	0.97	0.97
	ICA_E	1.01	0.99	1.01	1.01
	ICA_F	1.01	1.02	1.01	1.03
	ICA_G	0.99	0.99	0.97	0.97
	ICA_H	0.99	0.99	0.97	0.96

Supplemental Table S2. Additional Calibration Data and Associated Methods

Training Data	Model	Testing Data Imputation: missForest		Testing Data Imputation: NativeImputation	
		Unadjusted Calibration	Adjusted Calibration	Unadjusted Calibration	Adjusted Calibration
NativeImputation	Br_A	1.00	1.02	1.00	1.02
	Br_B	0.99	0.99	0.99	0.99
	Br_C	0.99	1.05	0.99	1.05
	Br_D	0.99	1.04	0.99	1.04
	Br_E	1.02	1.01	0.98	0.96
	CRC_A	1.01	1.03	1.01	1.03
	CRC_B	1.01	1.01	1.01	1.02
	CRC_C	1.01	1.04	1.01	1.04
	CRC_D	1.03	1.02	1.00	0.99
	CRC_E	1.02	1.01	1.00	0.98
	CRC_F	1.01	1.00	1.00	0.99
	CRC_G	1.01	1.00	1.00	0.98
	ICA_A	1.00	1.01	1.00	1.01
	ICA_B	1.00	1.01	1.01	1.01
	ICA_C	1.00	1.03	1.00	1.03
	ICA_D	1.01	1.00	0.99	0.99
	ICA_E	1.00	1.00	1.01	1.01
	ICA_F	1.00	1.02	1.00	1.02
	ICA_G	1.01	1.00	0.99	0.99
	ICA_H	1.01	1.00	1.00	0.99
MetaModel		Unadjusted Calibration	Adjusted Calibration		
	Lung Cancer	1.00	0.99		
	Breast Cancer	0.98	0.97		
	Colorectal Cancer	1.00	0.99		

Br\_=Breast Cancer Model; ICA\_=lung cancer model; CRC\_=colorectal cancer model.

Supplemental Table S2. (continued)

## SUPPLEMENTAL REFERENCES

References number 33 or lower are as cited in the main text.

34. H. OBRal. A random forests quantile classifier for class imbalanced data. Pattern Recognition 2019;90: 232-49.
35. Rahman MS, Ambler G, Choodari-Oskooei B, Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. BMC Med Res Methodol 2017;17(1):60.
36. "R Core Team RFfSC. R: A Language and Environment for Statistical Computing. 2019.





# 6

# CHAPTER

## THE VALUE OF ARTIFICIAL INTELLIGENCE IN LABORATORY MEDICINE – CURRENT OPINIONS AND BARRIERS TO IMPLEMENTATION

Ketan Paranjape\*

Michiel Schinkel\*

Richard D. Hammer

Bo Schouten

Rishi S. Nannan Panday

Paul W.G. Elbers

Mark H.H. Kramer

Prabath W.B. Nanayakkara

\* Both authors contributed equally to the manuscript

## ABSTRACT

### Background

As laboratory medicine continues to undergo digitalization and automation, clinical laboratorians will likely be confronted with the challenges associated with artificial intelligence (AI). Understanding what AI is good for, how to evaluate it, what are its limitations and how it can be implemented are not well understood. With a survey we aimed to evaluate the thoughts of stakeholders in laboratory medicine on the value of AI in the diagnostics space and identify anticipated challenges and solutions to introducing AI.

### Methods

We conducted a web-based survey on the use of AI with participants from Roche's Strategic Advisory Network that included key stakeholders in laboratory medicine.

### Results

In total, 128 of 302 stakeholders responded to the survey. Most of the participants were medical practitioners (26%) or laboratory managers (22%). AI is currently used in the organizations of 15.6%, while 66.4% felt they might use it in the future. Most had an unsure attitude on what they would need to adopt AI in the diagnostics space. High investment costs, lack of proven clinical benefits, number of decision makers, and privacy concerns were identified as barriers to adoption. Education in the value of AI, streamlined implementation and integration into existing workflows, and research to prove clinical utility were identified as solutions needed to mainstream AI in laboratory medicine.

### Conclusions

This survey demonstrates that specific knowledge of AI in the medical community is poor and that AI education is much needed. One strategy could be to implement new AI tools alongside existing tools.

### Impact Statement

Addressed a target population of participants in laboratory medicine who are key decision makers to either embrace AI or to refrain from implementing it. We identified key barriers and potential solutions to mainstream AI in laboratory medicine.

## INTRODUCTION

Advances in our understanding of biology, disease and molecular medicine have created a central role for laboratory medicine in the diagnostic work-up of many, if not most diseases. It is estimated that 70% of decisions regarding a patient's diagnosis, treatment, and discharge are in part based on results of laboratory tests (1). Unfortunately, the main cause of medical errors in the United States (US) is inaccurate diagnosis (2,3,4,5). The ever-increasing workload, high healthcare costs and need for improved precision, call for continuous optimization of the laboratory processes (6). With both healthcare and laboratory medicine (7) transitioning into an era of big data and artificial intelligence (AI), the ability to provide accurate, readily available and contextualized data is crucial. AI in healthcare is the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data generated from diagnostics, medical records, claims, clinical trials etc. AI algorithms can only properly function with reliable and accurate laboratory data (8). Automation and AI can fundamentally change the way medicine is practiced, as demonstrated by the recent applications in ophthalmology (9) and dermatology (10). Some possible applications of AI specific to laboratory medicine are presented in table 1. With the increasing role of laboratory medicine in many diseases, AI has the potential to improve diagnostics through more accurate detection of pathology, better lab workflows, improved decision support and reduced costs leading to higher efficiencies (8,11,12).

As laboratory medicine continues to undergo digitalization and automation, clinical laboratorians will likely be confronted with the challenges associated with evaluating, implementing, and validating AI algorithms, both inside and outside their laboratories. Understanding what AI is good for, where it can be applied, along with the state-of-the-art and limitations will be useful to practicing laboratory professionals and clinicians. On the other hand, the introduction of new technologies requires willingness to change the current structure and mindset towards these technologies, which are not always well understood. Historically, there has been resistance to the adoption of new technologies in the medical community (13).

With a web-based survey among stakeholders in laboratory medicine in the United States, we aimed to evaluate their current perspectives on the value of AI in the diagnostics space and identify anticipated challenges with the introduction of AI in this field, as well as resistance to introduction of this new technology in today's practice.

## AI IN LABORATORY MEDICINE

Today AI is occasionally used in laboratory medicine for enabling the effective use of resources, avoiding unnecessary tests, improving patient safety and alerting for abnormal results (14,15,16,17,18). AI is also being used in limited clinical usage for molecular/genomic testing (19,20,21) by accurately identifying variants and matching it to possible treatments.

## METHODS

### Survey development

A web-based survey on the use of AI in laboratory medicine was designed in several independent steps. First, 98 stakeholders participated in a two-week online discussion board on AI in diagnostics in August 2019. These participants were part of Roche's Strategic Advisory Network, a group consisting of laboratory medicine decision makers, practicing physicians and surgeons, point of care coordinators, anatomic pathologists, laboratory management, information technology management and senior leadership. Roche does not know the identity of the community members to protect their privacy. The online discussion board was moderated by C-space (Boston, Massachusetts) and developed questions to gain insights on potentially important topics to discuss in the survey. Open-ended as well as multiple choice questions were formulated based on content of the discussion board.

Next, two one-hour online group chats were organized on October 2<sup>nd</sup> and 3<sup>rd</sup> 2019 to discuss these questions and fine-tune their phrasing and to refine the answer possibilities to the multiple-choice questions. In these group chats, a total of 11 practitioners in laboratory medicine were asked to answer the initial survey questions one at the time, after which they could comment on each other's answers and discuss their opinions on and interpretations of the questions.

This thoroughly discussed survey was fielded to a group of 302 laboratory medicine practitioners that are part of Roche's Strategic Advisory Network (SAN) via email. These individuals were both available for completing surveys and are in a position to decide on embracing or refraining from using technologies like AI in their respective organizations. The survey was available from October 21<sup>st</sup> until November 1<sup>st</sup> 2019. The data was collected using a software platform called Conffirmit (Oslo, Norway) and all participants gave informed consent for their input to be used for research purposes.

Finally, as there are multiple different definitions of AI, for the sake of the survey we defined AI as - "Artificial intelligence (AI) in healthcare is the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data generated from diagnostics, medical records, claims, clinical trials etc. AI is truly the ability for computer algorithms to approximate conclusions without direct human input."

### Questions

In the survey we posed 21 questions that ranged from collecting demographic information to answering questions about if the respondents used AI in their organizations, what kind of improvements they would like to see in the current use of AI, how valuable they think AI will be in their practice and what challenges they feel exist. A full list of all the questions can be found in the appendix.

## Data analysis

### Quantitative data

Categorical data was analyzed using a Pearson chi-square test, considering a p-value of <0.05 to be statistically significant. The perceived value of AI was compared for different age groups and different experience levels with AI. The analyses were performed in Microsoft Excel (Microsoft, Redmond, Washington). Data from the multiple-choice questions were presented as percentages per category.

The number of participants that could be reached was highly dependent on the number of available advisors in the SAN network. With 302 available advisors and an acceptable response rate of above 40%, we got responses from 128 participants. The study was powered to detect a 20% difference across subgroups in how they valued AI.

### Qualitative data

An inductive approach<sup>22</sup> of direct content analysis was used to analyze the open-ended questions. Firstly, two researchers, BS (psychologist) and MS (internal medicine doctor), independently screened the answers and drafted a rough framework of themes. After consensus on the overarching themes, answers were independently coded with this framework by both researchers. Then conflicts were resolved by consensus to account for different interpretations of the answers. Coding was performed and bar charts were created in Excel (Microsoft, Redmond, Washington).

## RESULTS

### Demographics

The survey was fielded to 302 stakeholders in laboratory medicine, of whom 128 (42%) responded. The modal age group was aged between 41-50 (32.0%), while 23 (18.0%) of the respondents were younger than 41. The top 3 participants were physicians, laboratory managers and pathologists. See table 1 for further details on demographic information.

### Qualitative analysis

Based on the data, six main themes were derived (attitude, quality of care, organizational value, data analysis, prerequisites and education). The ‘attitude’ theme was further categorized into three subthemes (positive, unsure and negative). In order to prevent losing valuable information, multiple themes could be assigned to an answer. The specific content of these themes are presented in table 2, as well as in subsequent paragraphs, along with quotes from the survey participants. It should be noted that the attitude theme could be interpreted as being a separate sentiment analysis. However, this is not the case. Attitude was merely a theme in which many answers could be categorized according to both researchers.

In 173 of the 237 coded cases (73%) there was an initial agreement on the codes to be assigned. In 64 cases (27%) were there was a discrepancy between codes assigned by

**Table 1.** Baseline Characteristics of Survey Respondents (n = 128)

Characteristic	Number (percent) N=128
<b>Gender</b>	
Male	80 (62.5%)
Female	48 (37.5%)
<b>Age</b>	
31-40	23 (18.0%)
41-50	41 (32.0%)
51-60	32 (25.0%)
61-70	29 (22.7%)
70+	3 (0.2%)
<b>AI use</b>	
Currently use AI	20 (15.6%)
Not currently, may use AI in future	85 (66.4%)
Not currently and will never use AI	8 (6.3%)
Unsure about AI use	15 (11.7%)
<b>Role</b>	
Physicians	28 (22%)
Laboratory management	24 (19%)
Pathologists	21 (16%)
Executive-level management	16 (13%)
Purchasing/procurement management	5 (4%)
Information technology management	3 (2%)
Other	10 (8%)
<b>Employment Type</b>	
Hospital	38 (30%)
Other	26 (20%)
Academic medical center/Teaching hospitals	14 (11%)
Integrated health network	9 (7%)
Private clinics	7 (5%)
Physician lab offices, federal government acute care facility, Reference lab	13 (10%)

the two different coders, an extensive consensus procedure was followed. This resulted in a 100% agreement between coders after this consensus procedure.

### Current uses of AI in laboratory medicine

AI is currently used in the organizations of 20 (15.6%; 95%-CI: 9.8-23.1%) of the participants, while 85 (66.4%; 95%-CI: 57.5-74.5%) felt they might use it in the future and 8 (6.3%; 95%-CI: 2.7-11.9%) felt that they would never use AI. Respondents who use AI in their practice, use it for diagnosing diseases from images (30%), reviewing patients' risk profiles for certain conditions (40%), pre-empt rapid response solutions (30%) and for automatically releasing laboratory results and financial analytics (10%). Examples of specific use cases include AI to perform digital cell analysis for peripheral

**Table 2.** List of 6 themes derived from the survey

Theme	Examples of content
Attitude – Positive	Respondent showed a positive attitude towards AI rather than giving a really specific answer to the question.
Attitude – Negative	Respondent showed a negative attitude towards AI rather than giving a really specific answer to the question.
Attitude – Unsure	Respondent generally was not sure about the influence of AI in a certain area.
Quality of Care	Accessibility of care, accuracy of diagnoses, and early recognition of certain disease states.
Organizational Value	Providing quick results, reducing redundancy, and resource management.
Data-analysis	Analyzing large datasets (big data).
Prerequisites	Workable user interface, IT support, and better software.
Education	Education on specific tools, and on AI in general.

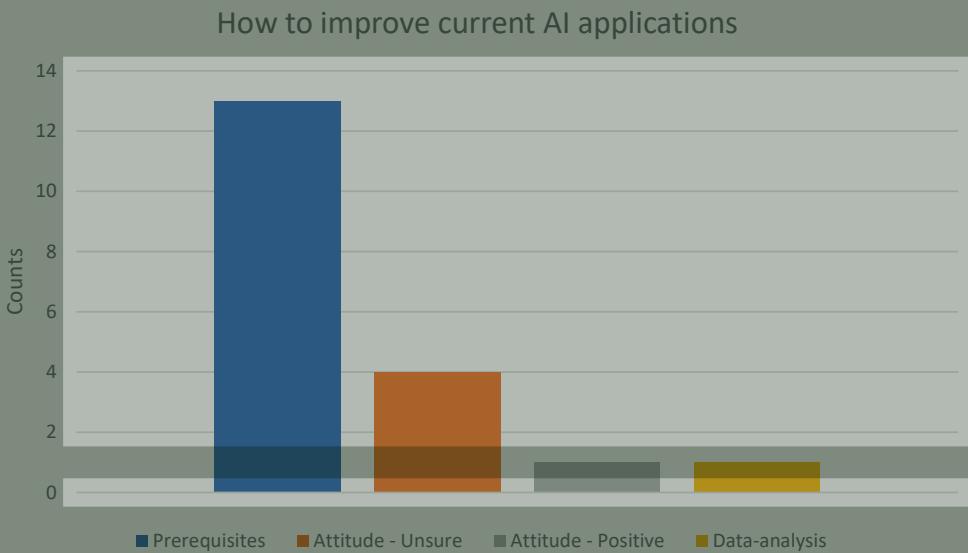
AI, artificial intelligence; IT, information technology.

bloods, analyse medical records laboratory data to determine which patients are at risk of infection, or to improve patients' outcomes and length of stays and re-admissions, pre-empt rapid response situations in hospitals and automated sepsis alerts to identify patients early.

When asked how respondents felt that these current applications could be improved, most of them answered that certain prerequisites (e.g. user interface, IT support, and better software) were needed. All 20 participants who currently use AI answered this question, see figure 1 for counts against themes. For example, respondent 119 answered: "*reduce the number of pop-ups in the EMR*" and respondent 70 said: "*We use AI for chatbots about common questions for diagnostics. The AI chatbots are not very intelligent. Need to make AI smarter*".

### Value of AI in practice

Regarding the potential use of AI in the diagnostics space, 90 (81%; 95%-CI: 72.6-87.9%) participants felt AI will be valuable in their organization within the next five years, of whom 20 (18%; 95%-CI: 11.4-26.5%) labelled it as expected to be extremely valuable. 21 (19%; 95%-CI: 12.1-27.5%) of the respondents felt like AI will not be valuable in their organization within the next five years. There was missing data on this question for 17 participants. To further examine whether the results were different in subgroups of respondents, we dichotomized the answers to finding AI valuable (including extremely valuable, very valuable and somewhat valuable; N=90) or not valuable at all (N=21). A chi-square test showed no significant difference in how participants in the different age groups valued AI in the diagnostics space ( $\chi^2=5.0947$  (4 degrees of freedom);  $p=0.28$ ). Also, there was no significant difference in how AI was valued between respondents who currently use AI in their practice (N=17), compared to respondents who have not



**Figure 1.** Answers to the survey question “How can current AI applications in your organization be improved? - categorized as counts per theme.

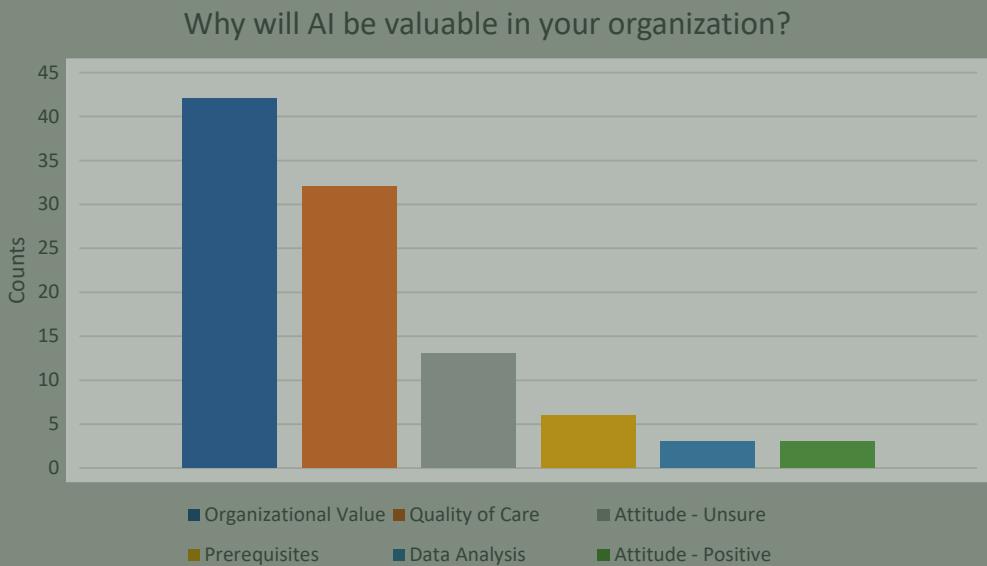
**Table 3.** Categorized subgroup results for finding AI valuable or not valuable in the diagnostics space. Percentage calculated row-wise.

Category	Subgroup	Valuable (n, %)	Not valuable (n, %)
<b>Age</b>	31-40	17 (85)	3 (15)
	41-50	22 (69)	10 (21)
	51-60	25 (83)	5 (17)
	61-70	23 (88)	3 (12)
	70+	3 (100)	0 (0)
<b>Experience</b>	Use AI	15 (88)	2 (12)
	Do not use AI	75 (80)	19 (20)

used it yet ( $\chi^2 = 0.6698$  (1 degree of freedom);  $p=0.41$ ). Table 3 shows the number of respondents to find AI valuable or not valuable in the different subgroups.

### Valuable

Respondents were also given the opportunity to elaborate on why they thought AI would or would not be valuable in their organizations. A plethora of reasons were given and all were coded by theme as shown in figure 2. Most answers indicated that AI could be valuable because of the ‘organizational value’ (e.g. quicker results, reduced redundancy, and resource management). As an example, respondent 87 answered: “it could make the lab more efficient by streamlining work flow”. Another frequently reported theme was ‘quality of care’ (e.g. accessibility of care, accuracy, and early recognition). This was



**Figure 2.** Answers to the survey question “Why will AI be valuable in your organization?” – categorized as counts per theme.

illustrated by respondent 100 who said “*might help in keeping patients informed of test results/appointments/follow up more efficiently*” and respondent 47 believed “*could have some useful clinical algorithms to identify problems before they are known by humans, but the technology is still in early development*”. Another substantial part of respondents, who thought AI would be valuable, were not sure about the reasons for this ('attitude – unsure'), as suggested by respondent 68: “*I'm not entirely sure, I just know something is there!*”.

### Not valuable

The 19% (95%-CI: 12.1-27.5%) of respondents who did not consider AI to be valuable in their organizations in the next 5 years, had more uniform responses. The answers were largely split between the themes ‘prerequisites’ (e.g. budget and strategic plan) and an unsure attitude. See figure 3 for more details. The missing prerequisites were for example presented by respondent 47: “*very expensive and we have very limited capital dollars that we need to use to refresh old technology*” and respondent 106 said “*it's not in our strategic plan to implement AI at this time*”. The unsure attitude towards AI was summed up by respondent 75: “*I'm not sure about the use of AI*”.

### Requirements for implementing AI

Participants were asked what they would need in order to feel comfortable with using AI in the diagnostics space. The majority of respondents had an “unsure attitude” towards what they needed most in order to adopt AI in their practice. For example, respondent 69



**Figure 3.** Answers to the survey question “Why will AI not be valuable in your organization?” – categorized as counts per theme.

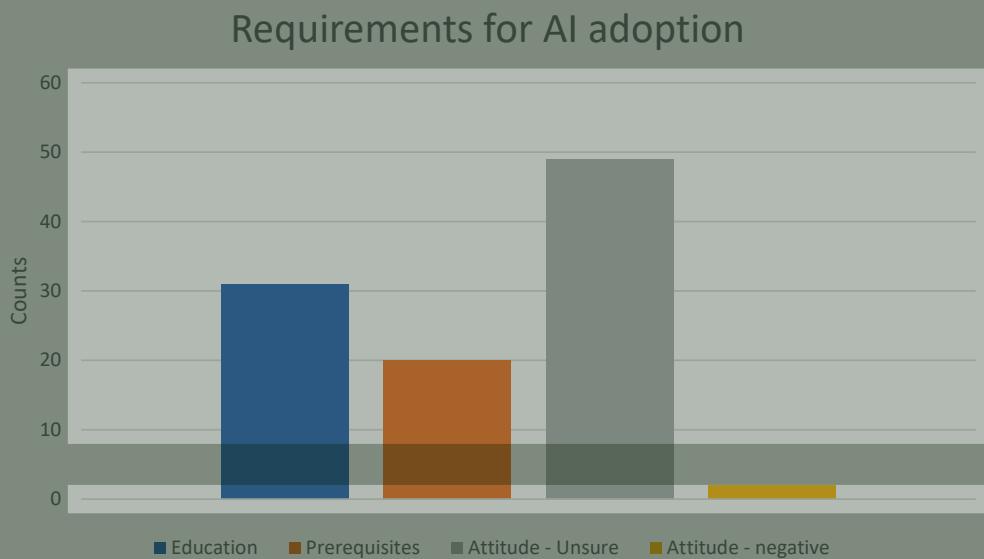
said: “*this seems like too forward thinking of a question that we aren’t yet prepared to answer*”. Others felt like they need education (e.g. specific to tools, and on AI in general). Respondent 40 answered: “*specific training to device*” and respondent 46 said: “*AI short course training*”. Most of the remaining group felt like they needed various prerequisites (e.g. support systems, certifications, and evidence of benefits) in order to feel comfortable to adopt AI in their practice. There was missing data for 23 of the participants, see figure 4 for the counts per theme.

In the next question, participants were asked to specify how they would like to be trained to use AI. 23 of 105 (22%; 95%-CI: 14.4-31.0%) participants felt that they were not able to speak to how they should be educated on these new technologies.

Finally, we asked participants to select persons within their organizations who they felt should be involved in the selection of AI equipment. Up to 10 individuals across an organization were identified that could be involved in evaluating a potential AI diagnostic solution. Respondent 56 said: “*Medical staff committees and physicians and mid-level providers that use AI, utilization review staff that monitor provider performance, IT department and leadership that maintain AI software.*”

## DISCUSSION

With this survey, we aimed to evaluate the thoughts of stakeholders in laboratory medicine on the value of AI in the diagnostics space and identify anticipated challenges with the introduction of AI in this field. About four in five respondents felt like AI will be valuable in their organization within the next five years, mostly because of organizational and patient oriented benefits. One in five respondents do not see any value in AI, which



**Figure 4.** Answers to the survey question: “what requirements are necessary in order for you to feel more comfortable with adopting AI?” – categorized as counts per theme

is often because of the lack of prerequisites like budget for the implementation and not being adopted as part of an overall strategic initiative. Maybe this is because of the bias management might have towards AI.

### The value of AI

The quantitative results in this survey are similar to those found in surveys on this subject amongst other healthcare professionals. Similar to our findings, surveys on AI for pathologists (23), medical students (24), physicians (25) and radiologists (26) all found that about 80% of participants feel like AI will be influential or valuable in their practice in the upcoming years. Interestingly enough, a 2019 survey shows that 84% of the general population in the US feels that AI will be at the centre of the next technological revolution (27).

All surveys on this subject in the medical community seem to show similar results regarding the perceived value of AI, independent of age or experience with AI in clinical practice. The fact that these results also overlap with the value of AI as perceived by the general population, raises concerns that specific knowledge on this subject has not yet penetrated the medical community at large and that the surveys on this subject just reflect the ongoing AI hype. Our survey adds to this concern by showing that many respondents are unsure about why AI would or would not be valuable, what is needed to comfortably adopt AI or how to be educated on AI.

There certainly seems to be a disconnect between the more positive views of information experts on AI and views of the medical community (28). In order to get all the benefits AI presents, while keeping its drawbacks to a minimum, drastic changes are needed in the medical community. There is a need for general AI training to the various

healthcare stakeholders as identified in a recent publication on the need to introduce AI training in medical education (29). In the meantime, training on new AI tools should also be the responsibility of the companies who provide the algorithms through extensive web-based training, along with on-site hands on training.

## Healthcare costs

Another highlight from this survey was the potential of AI to target high healthcare expenditure, since it can reduce and replace repetitive manual labour. Recent study has shown that AI can help reduce the waste in the US healthcare system in the range from \$760 billion to \$935 billion in 2019 (30). Respondents feel like AI can make the diagnostics process more efficient and decrease costs (1). For example, safely reducing the number of lab test ordered or the frequency of ordering repeat tests, is illustrated with a quote from respondent 94: *"Alert me to the fact that a lab test I ordered was already completed at another hospital system in the past week"*.

## Impact on jobs

Finally, we learned from the discussion board conducted prior to developing the survey that advisors also mentioned that they know AI is rapidly increasing in importance and value, and want it to evolve their roles, rather than replace them (31). *"AI will become an integral concept for healthcare. Whether diagnosis or process improvement in medicine, AI will impact the industry. For personalized medicine and improving diagnostic accuracy AI will drive decision making in the hands of providers"*, said respondent 12, an executive at a large integrated health network. Lab managers similarly feel that AI could create efficiencies that expedite their workflows, but want to ensure that they are still in control. Respondent 31 believes: *"AI needs to be used in the right spaces and not to eliminate med techs but to supplement them"*.

## Implementation strategies

In this survey, 19% of respondents did not see the added value of implementing AI in laboratory medicine, partially because of the high initial investment costs. This will be a limiting factor as long as the return on investment and clinical benefits of these tools are not well understood. A recent narrative review on the clinical applications of AI for sepsis validates this idea by identifying that a large gap that still remains between the development of AI algorithms and their clinical implementation (32). The question remains whether this gap between developing and clinical implementation might be caused by the resistance to implementing new technologies. Unfortunately, this can hold back research on this subject and thereby delay the gathering of evidence on whether AI tools can be beneficial and cost-effective in clinical practice on a large scale.

Although it was not mentioned in the survey itself, an interesting strategy to implement AI in laboratory medicine was discussed in one of the group chats that was used to shape the final survey. One of the participants disclosed that in their hospital, a new AI

tool was introduced alongside an existing tool that was used in routine clinical practice. The old tool was still used, but as a backup for when the AI tool failed. Practitioners were encouraged to try the AI tool, but could choose either of the available options. They gradually got familiar with the AI tool and could see the added value first hand. They ended up switching to the AI tool completely. This illustrates a viable way to integrate AI tools in healthcare. Although more expensive, this provides an opportunity to compare these tools in practice and allow the practitioners to feel comfortable with the tool before having to rely on it completely. See table 4 for our key recommendations.

### Patient view point

The overarching goal of implementing AI in clinical practice is to benefit the patient. Therefore, the patient's perspective should also be discussed. One of the respondents posed an interesting question in the online group chat: *"Should the patients be informed that some of the decisions are being recommended by AI?"* Another question is whether we should inform patients when an AI recommendation is not followed. Unfortunately, this burdens the patient as they now have to choose between the physician and the computer. Many algorithms are already being used in medicine, like the YEARS criteria (33) for pulmonary embolisms. Their role in the diagnostic process is rarely explained to or discussed with the patient, as only trained physicians can interpret the results of these algorithms. We therefore believe that a similar approach might be best when using more advanced algorithms, in which explainability and interpretation are an even larger problem. These tools and algorithms are an aid to complement the healthcare practitioners who are eventually responsible for the diagnostic process and decision making. Finally, from a provider's viewpoint they will need to know details of the algorithms they use to make decisions.

### Strengths of survey

We addressed a target population of participants who are currently in a position to influence organizational policies to either embrace new technologies or to refrain from using them in their laboratories. Any specific intervention to encourage the introduction

**Table 4.** Key recommendations for implementing AI in laboratory medicine

Area	Recommendation
Education	Need for general AI training in medical education – an approach has been proposed (29)
Implementation	Implement new AI tools alongside current tools, to give practitioners time to get comfortable and see benefits firsthand albeit only suggested by one respondent
Research	Research on AI in laboratory medicine should focus on generating clinical evidence of benefits and implementation

of AI in the diagnostics space should be tailored to such a population of decision makers. Another strength is that the results were independently analyzed by two researchers with different backgrounds, thereby minimizing the chance of interpretation bias. Finally, the questions were extensively scrutinized in the initial discussion board and group chats prior to fielding the final survey.

### **Limitations of survey**

The participants did not represent the entire population of practitioners who will be using AI in a diagnostic setting. We cannot generalize these findings to all laboratory medicine practitioners across multiple types of settings. Finally, the study population ( $N=128$ ) was relatively small for quantitative analyses, perhaps causing the non-significance of the chi-squared tests. Only a large difference in how AI was valued between groups would have shown significant results in the quantitative analysis.

## **CONCLUSION**

This survey shows that many stakeholders in laboratory medicine think that AI will be valuable to them in the near future, mostly given the 'organizational value' and expected improvements in 'quality of care', although vital prerequisites such as support systems, strategic plans and budgets need to be provided. The overall response to this and other similar surveys, raises the concern that specific knowledge on AI in the medical community at large is still poor. AI education in the medical community is much needed. As suggested by one respondent, one strategy to implement new AI tools could be to implement it alongside existing tools, so that practitioners can feel comfortable with the new tools and experience their added value in practice first hand, while awaiting further research studies on the clinical evidence, implementation and benefits of AI.

## REFERENCES

1. Badrick T. Evidence-based laboratory medicine. *Clin Biochem Rev.* 2013;34(2):43-46. <http://www.ncbi.nlm.nih.gov/pubmed/24151340>. Accessed December 22, 2019.
2. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ.* 2016;353. doi:10.1136/BMJ.I2139
3. *To Err Is Human.* To Err is Human: Building a Safer Health System Institute of Medicine (US) Committee on Quality of Health Care in America Linda T. Kohn, Janet M. Corrigan, Molla S. Donaldson , editors. Washington, D.C.: National Academies Press; 2000. doi:10.17226/9728
4. Brennan TA, Leape LL, Laird NM, et al. Incidence of Adverse Events and Negligence in Hospitalized Patients. *N Engl J Med.* 1991;324(6):370-376. doi:10.1056/NEJM199102073240604
5. Leape LL, Brennan TA, Laird N, et al. The Nature of Adverse Events in Hospitalized Patients. *N Engl J Med.* 1991;324(6):377-384. doi:10.1056/NEJM199102073240605
6. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. *Crit Rev Clin Lab Sci.* 2019;56(2):98-110. doi:10.1080/10408363.2018.1561640
7. Tizhoosh H, Pantanowitz L. Artificial intelligence and digital pathology: Challenges and opportunities. *J Pathol Inform.* 2018;9(1):38. doi:10.4103/jpi.jpi\_53\_18
8. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? *Clin Chem Lab Med.* 2018;56(4):516-524. doi:10.1515/cclm-2017-0287
9. 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
10. 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
11. Zhang Z. The role of big-data in clinical studies in laboratory medicine. *J Lab Precis Med.* 2017;2:34-34. doi:10.21037/jlpm.2017.06.07
12. Durant T. Machine Learning and Laboratory Medicine: Now and the Road Ahead. AACC. <https://www.aacc.org/publications/cln/articles/2019/march/machine-learning-and-laboratory-medicine-now-and-the-road-ahead>. Published 2019.
13. Safi S, Thiessen T, Schmailzl KJ. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR Res Protoc.* 2018;7(12):e11072. doi:10.2196/11072
14. Cismondi F, Celi LA, Fialho AS, et al. Reducing unnecessary lab testing in the ICU with artificial intelligence. *Int J Med Inform.* 2013;82(5):345-358. doi:10.1016/j.ijmedinf.2012.11.017
15. Gunčar G, Kukar M, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep.* 2018;8(1):411. doi:10.1038/s41598-017-18564-8
16. Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. *Clin Biochem.* 2019;69:1-7. doi:10.1016/J.CLINBIOCHEM.2019.04.013
17. Li L, Georgiou A, Vecellio E, et al. The Effect of Laboratory Testing on Emergency Department Length of Stay: A Multihospital Longitudinal Study Applying a Cross-classified Random-effect Modeling Approach. Wilber ST, ed. *Acad Emerg Med.* 2015;22(1):38-46. doi:10.1111/acem.12565
18. Garcia E, Kundu I, Ali A, Soles R. The American Society for Clinical Pathology's 2016-2017 Vacancy Survey of Medical Laboratories in the United States. *Am J Clin Pathol.* 2018;149(5):387-400. doi:10.1093/ajcp/aqy005

19. Xu J, Yang P, Xue S, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Hum Genet.* 2019;138(2):109-124. doi:10.1007/s00439-019-01970-5
20. Serag A, Ion-Margineanu A, Qureshi H, et al. Translational AI and Deep Learning in Diagnostic Pathology. *Front Med.* 2019;6:185. doi:10.3389/fmed.2019.00185
21. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* 2019;11(1):70. doi:10.1186/s13073-019-0689-8
22. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *Am J Eval.* 2006;27(2):237-246. doi:10.1177/1098214005283748
23. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *npj Digit Med.* 2019;2(1):28. doi:10.1038/s41746-019-0106-0
24. Pinto dos Santos D, Giese D, Brodehl S, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. *Eur Radiol.* 2019;29(4):1640-1646. doi:10.1007/s00330-018-5601-1
25. Oh S, Kim JH, Choi S-W, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. *J Med Internet Res.* 2019;21(3):e12422. doi:10.2196/12422
26. Collado-Mesa F, Alvarez E, Arheart K. The Role of Artificial Intelligence in Diagnostic Radiology: A Survey at a Single Radiology Residency Training Program. *J Am Coll Radiol.* 2018;15(12):1753-1757. doi:10.1016/j.jacr.2017.12.021
27. 2019 Artificial Intelligence Survey | Edelman. <https://www.edelman.com/research/2019-artificial-intelligence-survey>. Accessed December 22, 2019.
28. Bleas C, Bernstein MH, Gaab J, et al. Computerization and the future of primary care: A survey of general practitioners in the UK. *PLoS One.* 2018;13(12):e0207418. doi:10.1371/journal.pone.0207418
29. 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. doi:10.2196/16048
30. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System. *JAMA.* 2019;322(15):1501. doi:10.1001/jama.2019.13978
31. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.* 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
32. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. *Comput Biol Med.* October 2019;103488. doi:10.1016/J.COMBIOMED.2019.103488
33. van der Hulle T, Cheung WY, Kooij S, et al. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. *Lancet.* 2017;390(10091):289-297. doi:10.1016/S0140-6736(17)30885-1
34. Yuan C, Ming C, Chengjin H. UrineCART, a machine learning method for establishment of review rules based on UF-1000i flow cytometry and dipstick or reflectance photometer. *Clin Chem Lab Med.* 2012;50(12):2155-2161. doi:10.1515/cclm-2012-0272
35. Putin E, Mamoshina P, Aliper A, et al. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging (Albany NY).* 2017;8(5):1021-1033. doi:10.18632/aging.100968
36. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. *Big Data.* 2015;3(4):277-287. doi:10.1089/big.2015.0020

37. Nelson DW, Rudehill A, MacCallum RM, et al. Multivariate Outcome Prediction in Traumatic Brain Injury with Focus on Laboratory Values. *J Neurotrauma*. 2012;29(17):2613-2624. doi:10.1089/neu.2012.2468
38. Lin C, Karlson EW, Canhao H, et al. Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records. Zhu D, ed. *PLoS One*. 2013;8(8):e69932. doi:10.1371/journal.pone.0069932
39. Liu KE, Lo C-L, Hu Y-H. Improvement of Adequate Use of Warfarin for the Elderly Using Decision Tree-based Approaches. *Methods Inf Med*. 2014;53(01):47-53. doi:10.3414/ME13-01-0027
40. Gunčar G, Kukar M, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep*. 2018;8(1):411. doi:10.1038/s41598-017-18564-8

## APPENDIX

### List of survey questions

#### *Demographic Questions*

- **Single Select, Do not randomize**

First, we have a few demographic questions for you.

How do you identify in terms of **gender**?

Male

Female

Other (please specify):

Open End

- **Single Select, Do not randomize**

How **old** are you?

18-30

31-40

41-50

51-60

61-70

70+

- **Single Select, Do not randomize**

Which of the following describes your **organization's** interactions with **artificial intelligence (AI) as it relates to diagnostics**?

For those of you who haven't heard of Artificial Intelligence, "Artificial intelligence (AI) in healthcare is the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data generated from diagnostics, medical records, claims, clinical trials etc. AI is truly the ability for computer algorithms to approximate conclusions without direct human input."

Currently use AI	1
Do not currently use AI, but may in the future	2
Do not currently use AI, and never will in the future	3
I'm not sure about our use	4
Other (please specify):	5, Open End

## Survey Questions

### *Current Experience*

- **Multi Select, Randomize, Show if Q3=1**

You mentioned that your organization currently uses AI in relation to diagnostics. We have heard some of you use AI already for the following activities.

Which diagnostic-related activities does your organization use AI for? *Select all that apply.*

Diagnose diseases from imaging	
Review patients' risk profiles for certain conditions	
Pre-empt rapid response situations	
Produce automated alerts	
In addition, we also use AI for:	Anchor, Open End

- **Open Text, Show if Q3=1**

Thinking about how your organization **currently uses AI for diagnostics**, what could be improved?

### *Organization-wide challenges/solutions*

- **Open Text**

To get us started, we want you to think about how artificial intelligence could help you in your role. Think about what you would like AI to help you with, if you had no limits.

**Fill in the blank:** I wish there was artificial intelligence that could help me\_\_\_\_\_

- **Grid Rating**

We have previously heard from the community that **some of your organizations experience the following challenges.**

**How valuable** do you think AI will be in solving each of the following **challenges** in the next five years?

**Columns:**

Extremely valuable	4
Very valuable	3
Somewhat valuable	2
Not at all valuable	1

**Rows, Randomize:**

Independent hospitals and practices being acquired or merging	Code A
Physician groups consolidating to achieve scale and integration, and maintain independence from hospital systems	Code B
Closer alignment of health insurers and hospital/physician groups	Code C
Growth of alternative care delivery and management options that provide more consumer centric options	Code D
Healthcare systems more focused on reducing overall cost, eliminating redundancies and unwarranted care, and focusing on better management of health and wellness	Code E
Fee-for-value (risk based contracts) becoming increasingly prominent	Code F
Evolving population health management models that support movement towards total population health risk models with accountability for health management and total cost	Code G
Patient-driven healthy living and adoption of patient self-diagnosis/self-management tools and technology	Code H
Lab consolidations	Code I
Governmental legislation and policies	Code J

- **Open Text List, Require all, Show if Q7=Code A:J =4:2**

You mentioned that AI will **be valuable** in solving the following challenges.

***In what ways do you think AI will be valuable?***

- **Open Text List, Require all, Show if Q7=Code A:J=1**

You mentioned that AI **will NOT be valuable** in solving the following challenges.

**For what reasons do you think AI will NOT be valuable?**

**Diagnostic-specific solutions**

- **Single Select, Open End**

How valuable do you think AI will be for **your organization specifically related to diagnostics** in the next 5 years?

<b>Extremely valuable</b> , in this way:	4, Open End
<b>Very valuable</b> , in this way:	3, Open End
<b>Somewhat valuable</b> , in this way:	2, Open End
<b>Not at all valuable</b> , because:	1, Open End

- **Open Text List, Require 1**

What areas within the diagnostics function at your organization would **benefit from AI**?

In what ways would these areas benefit from AI?

1.	.....
2.	.....
3.	.....

- **Open Text List, Require 1**

What areas within the diagnostics function at your organization would **NOT benefit from AI**? In what ways would these areas NOT benefit from AI (and therefore, should be left alone?)

1.	.....
2.	.....
3.	.....

- **Grid Rating**

**How valuable** do you think AI will be to each of the following **diagnostic testing areas** in the next five years?

**Columns:**

Extremely valuable	4
Very valuable	3
Somewhat valuable	2
Not at all valuable	1

**Rows, Randomize:**

Clinical Chemistry	Code B
Immunoassay	Code C
Molecular	Code D
POC (Point of Care)	Code E
Tissue/Immunochemistry	Code G
Hematology & hemostasis	Code H
Urinalysis	Code I
Patient self-testing	Code J

- **Open Text List, Require all, Show if Q13=Code A:J =4:2**

You mentioned that AI will **be valuable** in the following diagnostic testing areas.

***In what ways do you think AI will be valuable?***

- **Open Text List, Require all, Show if Q13=Code A:J=1**

You mentioned that AI **will NOT be valuable** in the following diagnostic testing areas.

***For what reasons do you think AI will NOT be valuable in each of these areas?*****Evaluation/Buying Process**

- **Open Text List, Require 1**

**Who at your organization** do you think would be involved in the **selection** of diagnostic equipment/solutions **related to AI**?

Please list their **job titles** below.

1.

2.

3.

4.

5.

- **Multi Select, Randomize**

Which **departments at your organization** would be involved in decisions regarding the **selection** of diagnostic equipment/solutions **related to AI?** *Select all that apply.*

Administrators/Leadership

Finance

Sales

Marketing

Human Resources

IT

Legal

Compliance

Operations

Purchasing

Quality

Patient Advocacy

Department-specific (for ex: Hematology)

Other (please specify):

Anchor, Open End

Other (please specify):

Anchor, Open End

# 6

- **Open Text List, Not required**

Imagine you are advocating to use an AI tool in a specific diagnostic area at your organization.

What **potential challenges/pushback** (if any) do you think you would experience from each of the departments involved in the selection process? *Fill in details just for the departments for which you would expect challenges/pushback.*

---

Mask in answers selected in previous Q

---

## [Ethics/Education]

- **Open Text List, Require 1**

What **education/certification/additional requirements** would be necessary related to AI/diagnostics to make you/your organization feel more comfortable adopting AI in the future?

---

1.

---

2.

---

3.

---

- **Open Text List, Require 1**

Ideally, how would you/your immediate team be trained/educated on the use of AI related to diagnostics? What information would you like to learn?

---

1.

---

2.

---

3.

---

- **Single Select, Open End**

That's all of our questions today! Anything to add?

---

Yes, I'd like to add:

---

**Open End**

---

---

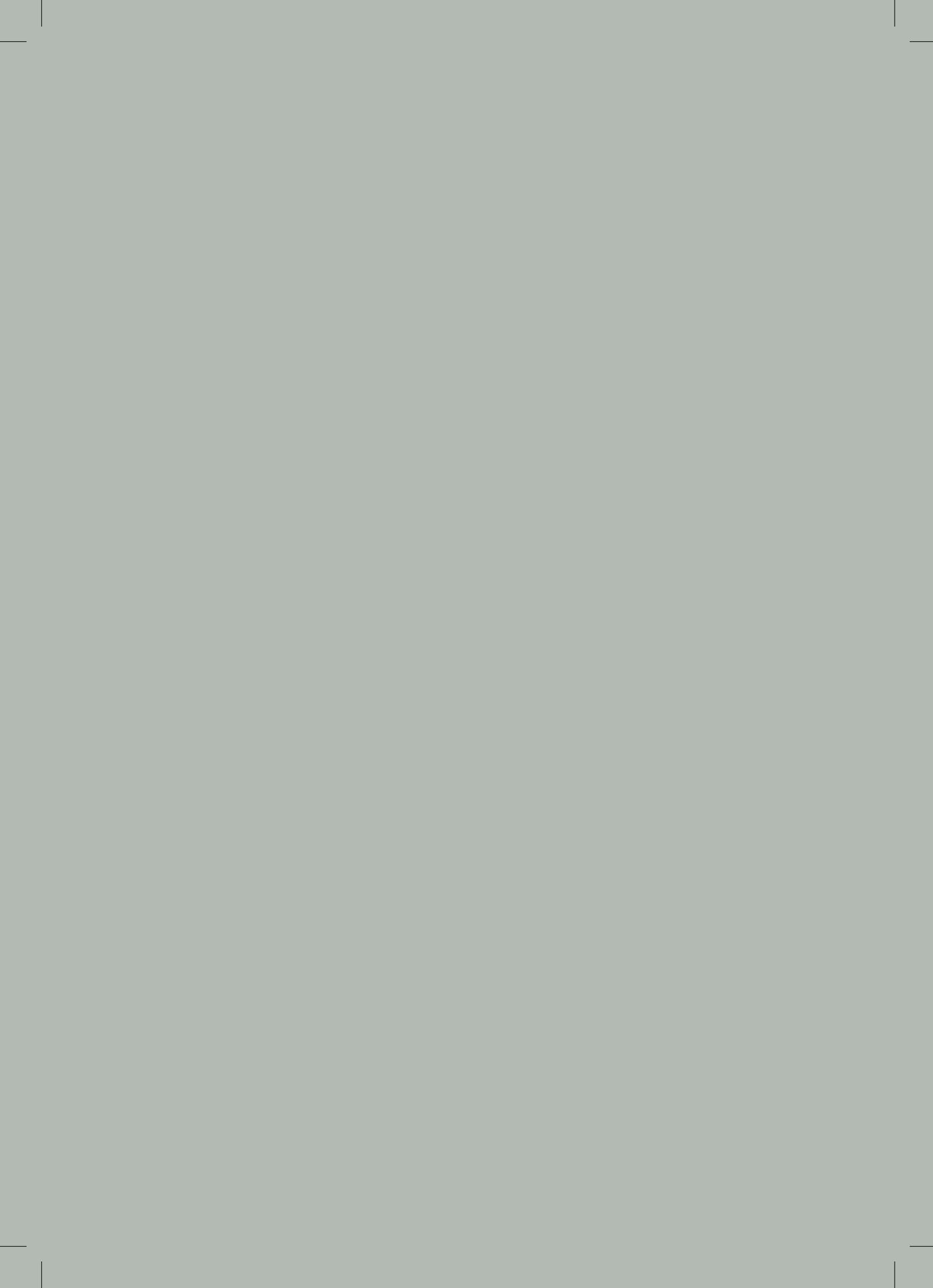
Nothing to add

---

## Summary of AI techniques used for specific laboratory applications\*

Test	AI technique applied	Outcomes
Discriminate between positive and negative urine tests <sup>34</sup>	Supervised Machine Learning (ML), Classification and Regression Tree (CART)	Better classifiers can reduce microscopic review rates by 30% and decrease significant losses in urinalysis
Predicting human age <sup>35</sup>	Deep neural networks	Found albumin concentration followed by glucose best identified age. Identified five markers (albumin, glucose, alkaline phosphate, urea and erythrocytes) as the most valuable to predicting human age
Predicting Type 2 diabetes <sup>36</sup>	Supervised ML, L1-regularized, logistical regression	Combined administrative claims, pharmacy records, healthcare utilization, lab results. Identified surrogate risk-factors such as chronic liver disease (odds ratio (OR) 3.71), High alanine aminotransferase (OR 2.26), esophageal reflux (OR 1.85), history of acute bronchitis (OR 1.45).
Predictor for traumatic brain injury (TBI) <sup>37</sup>	Logistic Regression, Relevance Vector Machine (RVM)	Predicted TBI outcome from lab data. Found Creatinine level was a clear predictor of outcome of traumatic brain injury. Glucose, albumin, osmolarity levels were also good predictors
Discover rheumatoid arthritis <sup>38</sup>	Linear kernel support vector machine, Natural Language Processing (NLP)	Combined clinical narratives and lab values from electronic health records (EHRs) to identify responders and non-responders for pharmacogenomics research
Warfarin adequacy, drug-drug interactions <sup>39</sup>	C4.5 decision tree, Random forest	Lab tests, alanine aminotransferase (ALT) and serum creatinine (SCr) combined with EHR data – warfarin dose, gender, age and weight. Automated results were “more accurate than clinical physicians’ subjective decision”
Hematological disease diagnosis <sup>40</sup>	Support vector machine (SVM), Naïve Bayesian Classifier, Random forest	Applying ML on laboratory blood test results can predict hematologic disease – prediction accuracies of 0.88 and 0.86 for five most likely diseases (multiple myeloma, amyloidosis, iron deficiency anemia, Purpura)

\*This table is included for readers to learn more about how specific AI techniques are used in laboratory medicine today.



# CHAPTER

7

## INTRODUCING ARTIFICIAL INTELLIGENCE TRAINING IN MEDICAL EDUCATION

Ketan Paranjape\*

Michiel Schinkel\*

Rishi S. Nannan Panday

Josip Car

Prabath W.B. Nanayakkara

\*Both authors contributed equally to the manuscript

## 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.

7

### 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 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).

**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 health 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.

## 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 biopsychosocial 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

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

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

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 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 <sup>a</sup>	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 <sup>b</sup> , AI <sup>c</sup> fundamentals, Ethics and Legal	Data sets <ul style="list-style-type: none"> <li>• HealthData.gov (99)</li> <li>• Public datasets in health care (100)</li> <li>• University of California San Francisco Data Resources (101)</li> </ul> AI fundamentals <ul style="list-style-type: none"> <li>• AI 101 course from MIT<sup>d</sup> (102)</li> </ul> Ethics, Law <ul style="list-style-type: none"> <li>• Teaching AI, Ethics, Law and Policy (103)</li> <li>• AI Law (104)</li> </ul> EHR Training (105)
Medical School – Clinical Phase	Familiarize with AI based clinical applications, Expand knowledge beyond basic principles of data/AI	Clinical Utility <ul style="list-style-type: none"> <li>• Overview of Clinical applications of AI (106)</li> <li>• AI for Health and Health Care (US Department of Health and Human Services) (107)</li> </ul> Center for AI in Medicine and Imaging (108) AI in Health care Accelerated Program (109)
USMLE <sup>e</sup>	Introduce questions on data sciences, AI, working with EHRs	Data Science Courses (110,111,112)
Residents	Detailed knowledge on clinical applications, Table 4 Attend conference in health care AI	
Specialist	Stay up to date on Data/AI through CME <sup>f</sup> credits, Attend conference in health care AI	Table 2, Table 4

<sup>a</sup>MCAT: Medical College Admission Test. <sup>b</sup>EHRs: electronic health records. <sup>c</sup>AI: artificial intelligence. <sup>d</sup>MIT: Massachusetts Institute of Technology. <sup>e</sup>USMLE: United States Medical Licensing Examinations. <sup>f</sup>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, Abbaftati 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](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.iclbezp1.cc.ic.ac.uk/stable/2251299?seq=1#metadata\\_info\\_tab\\_contents](https://www-jstor-org.iclbezp1.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. Alexović 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/](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, Halama 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/NEJMmp1606181.
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, Scherbier AJA. 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

# 8

## IMPLEMENTING ARTIFICIAL INTELLIGENCE IN HEALTH CARE: DATA AND ALGORITHM CHALLENGES AND POLICY CONSIDERATIONS

Ketan Paranjape  
Michiel Schinkel  
Paul W.G. Elbers  
Mark H.H. Kramer  
Prabath W.B. Nanayakkara

## ABSTRACT

### Background

Artificial Intelligence (AI) is already driving fundamental changes in health care operations and patient care, and is showing promise to significantly advance the quadruple aim - enhance the patient experience, improve population health, reduce costs and improve the provider experience. However, along with these impressive advances and the realistic potential to transform health care in the near term come thorny questions about liability and accountability, algorithmic bias and representative data, and the ability to accurately interpret and explain data.

We address the challenges of implementing AI in routine health care practice by looking at the role of data and algorithms and the implications for medical malpractice. We then summarize ongoing efforts to create AI policy and regulation globally to address these challenges in order to enable the mainstreaming of AI in health care.

### Methods

A systematic literature search was performed searching the electronic databases PubMed, and IEEE from inception to February 2019. Search terms related to AI were combined with terms regarding policy, legal, regulation, and law.

### Conclusion

AI has the potential to drive valuable transformation in health and in the health care ecosystem. However, several concerns continue to impede the assimilation of AI into the mainstream of health care and multiple other fields. These concerns include algorithmic transparency, liability, accountability, algorithmic bias, representative data, interpretability and explainability.

Numerous governments, consortiums and academic or scientific groups have assembled expert stakeholders from multiple disciplines to work toward recommending action steps to alleviate these concerns. A critical component of these recommendations is making long-term investments in AI research along with understanding and addressing the ethical, legal and societal implications of AI. Additional steps include creating awareness of biases and potential harm along with explaining the procedures used by an algorithm and how a decision was made. Establishing independent councils to provide impartial advice to the government, the private sector and the public on topics related to use of algorithms will also help alleviate concerns with taking AI mainstream.

For health care professionals, the recommendations by the American Medical Association (AMA) may hold the most appeal, as they advance the concept that physicians and a machine working in combination have the greatest potential to improve clinical decision-making and patient health outcomes.

Finally, this field is still evolving and thus the entire industry may have to take a wait and see approach before settling on the right set of policies and regulations to mainstream AI in health care.

## **Keywords**

Artificial intelligence, Policy, Algorithm, Transparency



## INTRODUCTION

Artificial intelligence (AI) in health care (1) is the use of a collection of computing technologies that perform perception, learning, reasoning, and decision-making tasks that are designed to augment human intelligence.

The goals of AI and its applications should be to advance the quadruple aim in health care (2,3) – enhance the patient experience, improve population health, reduce costs and improve the provider experience. AI in health care (4) can be used to augment the diagnostic process, automate image interpretation, improve clinical trial participation, reduce medication errors due to wrong dosage, detect insurance fraud, and improve the physician's workflow by providing point of care learning, clinical documentation and quality-measurement reporting (5).

As we digitize health care (6), we must ensure AI augments and enhances human capabilities and interactions, with the goal of building trust and avoiding failures. For example, misclassification of a malignant tumor can be costly and dangerous, and a wrong diagnosis can have far-reaching consequences for a patient, such as having further investigations and interventions recommended or even being sent to hospice.

The responsibility of establishing guidelines and policies lies under the aegis of government and regulating bodies, ideally with input from key stakeholders that include physicians, administrators, public and private institutions and patients. Along with developing new and retooled laws and regulations, these bodies should design policies to encourage helpful innovation, generate and transfer expertise, and foster broad corporate and civic responsibility to address critical issues raised by AI technologies.

As AI systems become mainstream in health care, they are susceptible to errors and failures. In a recent study an AI system was trained to learn which patients with pneumonia had a higher risk of death. It inadvertently classified patients with asthma as being at lower risk. This was because the system failed to comprehend that people with pneumonia and a history of asthma were directly admitted into the hospital and received treatment that significantly reduced their risk of dying. The machine therefore interpreted this as someone with pneumonia and asthma having a lower risk of death (7).

Another potential obstacle to AI going mainstream is the ability to “bias (8)” the data to cause misdiagnosis. Adversarial attacks are the ability to misclassify an output by engineering the inputs. Areas like dermatology, ophthalmology and radiology are susceptible as there are enormous incentives from which providers could benefit (9). For example, per CMS guidelines, insurance companies have to pay for vitrectomy surgery for a confirmed diabetic retinopathy diagnosis. In order to reduce the number of procedures without making a change to the CMS policy, could an insurance company use adversarial noise to bias the positive images?

Ultimately, the impact on end-users depends on how they perceive, interpret and tolerate these shortcomings. As AI becomes more embedded in our daily lives, mistakes could have serious if not deadly consequences. Whether the mistake is an accident caused by a self-driving car (10) or a misdiagnosis of a health condition, AI systems are and will increasingly be put under heavy scrutiny.

In this paper, we will address some challenges of implementing AI in routine health care practice by looking at the role of data and algorithms and the implications for medical malpractice. We will then share ongoing efforts in creating AI policy and regulation to address these challenges in order to enable the mainstreaming/mass adoption of AI in health care.

## CHALLENGES FOR ARTIFICIAL INTELLIGENCE IN HEALTH CARE

AI is becoming an increasingly advanced, sophisticated, and meaningful field, and its uses and implications are far-reaching. For example, researchers at Stanford recently developed an algorithm that can detect pneumonia from chest x-rays at a level exceeding the ability of practicing radiologists (11). Scientists at Google also developed a deep learning approach that could predict inpatient mortality, unexpected readmissions, and long length of stay more accurately than existing models by mining data from the electronic health records (12).

At the heart of AI's value proposition is the ability to process vast amounts of data and then act on that data through algorithms using techniques such as machine learning and cognitive computing (13). Harnessing the power of machines to process data much faster than any human being can, AI technologies have the potential to identify health care diagnoses, treatment plans and trends quickly by sifting through information and analyzing patient histories. These findings could support, inform, and enable physician decision-making.

Two factors that present challenges to implementing AI in health care are the role of algorithms and data and the potential ramifications for medical malpractice.

### Role of Algorithms, Data

Computer algorithms are at the core of AI and are part of applications in education, financial services, health care, navigation, and manufacturing. These algorithms are being used to make health care decisions like prioritizing activities for staff members or triggering interventions for admitted patients—such as was reported at John Hopkins Hospital (14). In these situations, how can one trust the algorithm to do the right thing, every time? Take the case of 'Deep Patient' where researchers at Mount Sinai Hospital applied deep learning to 700,000 patient records. Without expert guidance the tool was able to identify patterns and predict the onset of diseases such as cancer of the liver. The tool also predicted the onset of schizophrenia but offered no clue as to how it did so. For a condition that is very difficult to predict even by the most experienced psychiatrist, the way the AI system came up with its decision is known as the "black box" problem (15). So how can we trust such a system?

Even more concerning, what are the consequences for the patient? If physicians tell their patients they are going to develop schizophrenia in the future and over time they do not develop it, what impacts does this have on the patients? This raises another ethical dilemma as well: when physicians realize they made a wrong diagnosis due to AI, do they have an obligation to tell their patients about it?

Decisions made by predictive algorithms can be obscure because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations).

In considering the application of algorithms for individual patient treatment decisions, it should be noted that the physician will have to interpret the prediction model, as it is created by running the same algorithm on a mass scale. The physician will have to decide that the prediction score is reliable and accordingly propose a diagnosis.

Along with algorithms, AI relies heavily on patient data. To train a machine to identify specific conditions, hundreds of thousands of data elements are needed. For example, Google is using 46 billion data points to predict the medical outcomes of hospital patients (16). The challenge here is that this need for data runs up against current models of patient privacy, consent, and control. Implementing AI in health care could violate HIPAA (Health Insurance Portability and Accountability Act) policies (17).

Although HIPAA does not specifically address technologies such as AI, HIPAA was put in place to protect individuals' medical records and other personal health information. And there is a rapidly emerging move by consumers and patients to stop the "gold rush" mentality of industry to toss aside privacy and consent models (18). Companies like Luna DNA and other blockchain formulations are empowering consumers and patients to share data (19), but retain control.

Patients are now understanding the importance of meaningful notice, consent, and control over their data, both HIPAA-covered data and data outside of HIPAA jurisdiction.

## Implications for Medical Malpractice

As algorithms ingest large quantities of high-quality datasets from across the health care ecosystem, the use of AI will, over time, result in fewer misdiagnoses and errors (20). The AI machine will be able to predict diagnosis based on complex relationships between the patient and expected treatment results without explicitly identifying or understanding those connections. We are now entering an era where the medical decision-making burden shifts from the physician to an algorithm. What happens when the physician pursues an improper treatment—based on an algorithm—that results in an error?

In the US, medical malpractice is a professional tort system that holds physicians liable when the care they provide to patients deviates from accepted standards so much as to constitute negligence or recklessness (21). The system has evolved around the conception of the physician as the trusted expert, and presumes for the most part that the diagnosing or treating physician is entirely responsible for his or her decisions and thus accountable if the care provided is negligent or reckless. In England and Wales, medical liability is under the law of tort, specifically negligence (22). It is general practice in cases of clinical negligence that National Health Service Trusts (NHS) and Health Authorities are liable and are the bodies that are sued, rather than individual clinicians. The NHS has provided guidance that it will accept full financial liability where negligent harm has occurred, and will not seek to recover costs from the health care professional involved (23).

Who is liable for erroneous care based on a decision made by a machine is a hard problem to solve. Future malpractice guidelines should incorporate such considerations by including health care professionals but also software companies that created the algorithm.

In 2018, IDx-DR became the first and only FDA-authorized AI system used for the autonomous detection of diabetic retinopathy using deep learning algorithms (24). IDx, the company that developed the system, carries medical malpractice and liability insurance (25). The autonomous diagnostic AI is responsible for performing within specification for on-label use of the device, while in an off-label situation, the liability for an erroneous treatment decision typically would be with the physician using it off-label.

As AI evolves rapidly, the adoption and diversity of applications in medical practice are outpacing the critical need of establishing standards and guidelines to protect the health care community. There is a need for key stakeholders from both the public and private sector to collaborate and recommend policy guidelines to enable the safe use of AI.

## ONGOING EFFORTS IN CREATING AI POLICY

A systematic search revealed multiple initiatives underway in governments, academia, international consortiums, organizations and technology companies where policy guidelines and recommendations are being identified to address both the data and algorithm components of AI. Subject matter experts in these groups include members from science, engineering, economics, ethics, regulation and policy. While the majority of these recommendations do not directly address the health care industry, many could be applicable to the use of AI in health care.

In the following table we briefly summarize activities, in chronological order starting in 2016, that have gained significant momentum based on their impact and highlight if these are guidelines and recommendations to drive policy or enacted into legislation. We also call out if these recommendations apply to health care.

**Table 1.** Summary of global initiatives developing AI policy guidelines

#	Initiative	Summary
1.	General Data Protection Regulation (2016; GDPR) (26)	<ul style="list-style-type: none"> <li>GDPR explicitly addresses algorithmic discrimination by –           <ul style="list-style-type: none"> <li>“Data Sanitization” – removing special categories from data sets used in automated decision making.</li> <li>“Right to Explanation” whereby data subjects are entitled to “meaningful information about the logic involved, as well as the significance and the envisaged consequences” when automated decision making or profiling takes place</li> </ul> </li> <li>Enacted into legislation</li> </ul>
2.	Group of Seven (2016; G-7) (27)	<ul style="list-style-type: none"> <li>G-7 ministerial meeting on information and communication technology proposed setting up an international set of basic rules for developing AI.</li> </ul>

**Table 1.** (continued)

#	Initiative	Summary
3.	The One Hundred Year Study on Artificial Intelligence (2016; US) (28)	<ul style="list-style-type: none"> <li>• Study by Stanford University to address concerns about the individual and societal implications of AI</li> <li>• Three general policy recommendations that include accruing technical expertise in AI, removing perceptions and impediments to AI research and increasing funding for interdisciplinary studies</li> </ul>
4.	National Science and Technology Council (2016; US) (29)	<ul style="list-style-type: none"> <li>• NTSC subcommittee on Machine Learning and Artificial Intelligence</li> <li>• Strategy to coordinate all US government activities in AI.</li> <li>• Seven-point strategy with recommendations specific to health care (explainability and transparency)</li> </ul>
5.	Royal Statistical Society (RSS) (2017; UK) (30)	<ul style="list-style-type: none"> <li>• RSS's recommendations to the House of Commons Science and Technology Select Committee inquiry around the use of algorithms in decision-making</li> <li>• Recommendations include setting up an independent data ethics council to provide advice to government, public and private sector on the use of algorithms</li> </ul>
6.	Association of Computing Machinery (ACM) (2017; US, Europe) (31)	<ul style="list-style-type: none"> <li>• Developed by ACM's US Public Policy Council and Europe Council Policy Committee</li> <li>• Seven principles to address potential harmful biases generated by algorithms - awareness, access, accountability, explanation, data provenance, auditability and validation</li> </ul>
7.	Organisation of Economic Co-Operation and Development (2017; OECD) (32).	<ul style="list-style-type: none"> <li>• OECD's report on Algorithms and Collusion recommended that policy approaches should be developed in co-operation with competition law enforcers, consumer protection authorities, data protection agencies, relevant sectorial regulators and organizations of computer sciences with expertise in deep learning</li> </ul>
8.	AMA Guidance for Health Care Stakeholders (2018; US) (33)	<ul style="list-style-type: none"> <li>• First policy on health care Augmented Intelligence</li> <li>• Report provided that the overarching goal of AI in health care is to be human-centered and augment human intelligence and advance the quadruple aim: improve population health; improve health outcomes and patient satisfaction; increase value; and improve health care team satisfaction</li> </ul>
9.	International Telecommunications Union (ITU) and World Health Organization (WHO) Joint Focus Group on Artificial Intelligence for Health (34)	Recently formed focus group to identify opportunities for international standardization of AI for health-relevant data, information, algorithms, and processes, which will foster the application of AI to health issues on a global scale. The goal is to establish a standardized assessment framework with open benchmarks for the evaluation of AI-based methods for health, such as AI-based diagnosis, triage or treatment decisions.

## General Data Protection Regulation (GDPR)

The GDPR (26), adopted by the European Parliament and Council in April 2016, is the first piece of EU legislature that explicitly addresses algorithmic discrimination. Recital 71 states a requirement to "implement technical and organizational measures that prevent, inter alia, discriminatory effects on natural persons based on racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual

orientation, or that result in measures having such an effect.” These characteristics are referred to as “special categories”, and have their basis in non-discrimination legislation, such as *Article 14* of the European Convention on Human Rights (EC, 2010).

Moving beyond recitals (35), GDPR addresses algorithmic discrimination by two key principles. The first, *data sanitization*, is the removal of special categories from datasets used in automated decision making. This principle is introduced by *Article 9: Processing of special categories of personal data*, which establishes a *prima facie* prohibition against “the processing of data revealing racial or ethnic origin” and other “special categories”. It is strengthened under *Article 22: Automated individual decision-making, including profiling*, which specifically prohibits “a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her” that is “based on the special categories of personal data referred to in Article 9” (art. 22(2)).

The second principle, *algorithm transparency*, introduces the “right to explanation” (30), whereby data subjects are entitled to “meaningful information about the logic involved, as well as the significance and the envisaged consequences” when automated decision making or profiling takes place (art. 13(2)(f); art. 14(2)(g)). In *Article 12: Transparent information, communications and modalities for the exercise of the rights of the data subject*, the GDPR further specifies that such information must be provided “in a concise, transparent, intelligible and easily accessible form, using clear and plain language.”

## G-7

At the April 2016 Group of Seven (G-7) ministerial meeting on information and communication technology in Shikoku, Japan’s communication minister Sanae Takaichi proposed setting up an international set of basic rules for developing AI (27). The guidelines will include AI programs to be designed to not pose dangers to human lives and physical safety, have emergency stops and correct actions in real-time, protect themselves against cyber-attacks so that people with malicious intentions cannot take control and be transparent to examine AI’s actions (36).

## The One Hundred Year Study on Artificial Intelligence

In the fall of 2014 a long-term investigation of the field of AI and its influence on people, their communities, and society called “The One Hundred Year Study on Artificial Intelligence” was commissioned (28). The study considered the science, engineering, and deployment of AI-enabled computing systems. The study panel reviewed AI’s progress in the recent years, envisioned the potential advances that lay ahead, and described the technical and societal challenges and opportunities the field raised in areas of ethics, economics, and the design of systems compatible with human cognition.

To help address the concerns about the individual and societal implications of rapidly evolving AI technologies, the study panel offers three general policy recommendations (28). The first is to define a path toward accruing technical expertise in AI at all levels of government. Effective governance requires more experts who understand and can analyze

the interactions between AI technologies, programmatic objectives, and overall societal values. The second focuses on removing the perceived and actual impediments to research on the fairness, security, privacy, and social impacts of AI systems. The third recommends increasing public and private funding for interdisciplinary studies of the societal impacts of AI.

### National Science and Technology Council

To coordinate all federal activities in AI, in 2016 the United States government formed a National Science and Technology Council (NSTC) subcommittee on Machine Learning and Artificial Intelligence (37). The NSTC then directed a subcommittee on Networking and Information Technology Research and Development (NITRD) to create a National Artificial Intelligence Research and Development Strategic Plan (38).

The plan laid out a seven-point strategy that included:

- Making long-term investments in AI research
- Developing effective methods for human-AI collaboration
- Understanding and addressing the ethical, legal, and societal implications of AI
- Ensuring the safety and security of AI systems
- Developing shared public datasets and environments for AI training and testing
- Measuring and evaluating AI technologies through standards and benchmarks
- Better understanding the national AI R&D workforce needs (29)

For health care, the recommendation was to improve explainability and transparency related to the use of AI algorithms, as they are based on deep learning and are opaque to users, with few existing mechanisms for explaining their results. From a liability perspective, physicians need to know why a decision was suggested and need explanations to justify a diagnosis or a course of treatment.

### Royal Statistical Society (RSS)

In April 2017, the Royal Statistical Society (UK) (30) provided its recommendations to the House of Commons Science and Technology Select Committee inquiry for the use of algorithms in decision-making. The inquiry was focused on issues of governance, transparency, and fairness relating to algorithms.

The recommendations included the establishment of an independent Data Ethics Council which could provide impartial advice to the government, the private sector and the public on topics related to use of algorithms, allow existing laws (e.g., anti-discrimination) to manage issues arising due to the unfair use or results from using algorithms, develop professional standards for data science (including the incorporation of strong ethical training in data science courses), and making use of existing industry regulators to take on monitoring of outcomes from algorithms to check for bias.

## Association of Computing Machinery (ACM)

In May 2017, the ACM US Public Policy Council and the ACM Europe Council Policy Committee issued a set of seven principles designed to address: (a) potential harmful biases generated by algorithms that are not transparent; and (b) biased input data used to train these algorithms (31). These recommendations were not health care specific, but laid the foundation to provide context for what algorithms are, how they make decisions, and the technical challenges and opportunities to prevent and mitigate potential harmful bias.

The seven principles for algorithmic transparency and accountability focused on creating *awareness* of biases and potential harm, *access and redress* for individuals that are affected, *accountability* for using algorithms to make decisions, *explanation* of the procedures followed by the algorithm and decision made, *data provenance*, *auditability* of models, algorithms, data and decision and finally *validation and testing* of methods and results.

## Organisation for Economic Co-operation and Development (OECD)

In September 2017, OECD published a report on Algorithms and Collusion (32). The paper addressed challenges algorithms present for both competition law enforcement and market regulation. It also reported on market regulation and how traditional tools might be used to tackle forms of algorithmic collusion. Given the multi-dimensional nature of algorithms, the report suggests that policy approaches should be developed in cooperation with competition law enforcers, consumer protection authorities, data protection agencies, relevant sectorial regulators and organizations of computer sciences with expertise in deep learning.

## American Medical Association (AMA) Guidance for Health Care Stakeholders

In Jun 2018, the AMA's House of Delegates comprised of proportional representations of every major national medical specialty society and state medical associations adopted its first policy on health care Augmented Intelligence (33). The report accompanying the policy included a discussion of current generation AI systems that should augment clinical decision-making of physicians. It advanced that concept that physicians and machine working in combination improve clinical decision-making and patient health outcomes. In order to underscore the central role that humans must continue to play in health care even when enhanced with AI systems, the report utilized the term augmented intelligence instead of artificial intelligence. This reflected terminology that was utilized by others deploying AI systems in health care (such as IBM, Microsoft and Siemens). The policy report stated that the overarching goal of AI in health care is to be human-centered and augment human intelligence and advance the quadruple aim: improve population health; improve health outcomes and patient satisfaction; increase value; and improve health care team satisfaction.

**Table 2.** Summary of AMA Policy Recommendations

#	<b>Summary of AMA Policy Recommendations</b>
1	Leverage its ongoing engagement in digital health and other priority areas for improving patient outcomes and physicians' professional satisfaction to help set priorities for health care AI.
2	Identify opportunities to integrate the perspective of practicing physicians into the development, design, validation and implementation of health care AI.
3	Promote development of thoughtfully designed, high-quality, clinically validated health care AI that: a. is designed and evaluated in keeping with best practices in user-centered design, particularly for physicians and other members of the health care team; b. is transparent; c. conforms to leading standards for reproducibility; d. identifies and takes steps to address bias and avoids introducing or exacerbating health care disparities including when testing or deploying new AI tools on vulnerable populations; and e. safeguards patients' and other individuals' privacy interests and preserves the security and integrity of personal information.
4	Encourage education for patients, physicians, medical students, other health care professionals, and health administrators to promote greater understanding of the promise and limitations of health care AI.
5	Explore the legal implications of health care AI, such as issues of liability or intellectual property, and advocate for appropriate professional and governmental oversight for safe, effective, and equitable use of and access to health care AI.

## **International Telecommunications Union (ITU) and World Health Organization (WHO) Joint Focus Group on Artificial Intelligence for Health**

The International Telecommunication Union (ITU) has established a new Focus Group on "Artificial Intelligence for Health" (FG-AI4H) in partnership with the World Health Organization (WHO) (34). FG-AI4H will identify opportunities for international standardization of AI for health-relevant data, information, algorithms, and processes, which will foster the application of AI to health issues on a global scale. In particular, it will establish a standardized assessment framework with open benchmarks for the evaluation of AI-based methods for health, such as AI-based diagnosis, triage or treatment decisions.

## **SUMMARY AND CONCLUSION**

AI has the potential to drive valuable transformation in health and in the health care ecosystem. However, several concerns continue to impede the assimilation of AI into the mainstream of health care and multiple other fields. These concerns include algorithmic transparency, liability, accountability, algorithmic bias, representative data, interpretability and explainability.

Numerous governments, consortiums and academic or scientific groups have assembled expert stakeholders from multiple disciplines to work toward recommending action steps to alleviate these concerns. A critical component of these recommendations is making long-term investments in AI research along with understanding and addressing

the ethical, legal and societal implications of AI. Additional steps include creating awareness of biases and potential harm along with explaining the procedures used by an algorithm and how a decision was made. Establishing independent councils to provide impartial advice to the government, the private sector and the public on topics related to use of algorithms will also help alleviate concerns with taking AI mainstream.

For health care professionals, the recommendations by the American Medical Association (AMA) may hold the most appeal, as they advance the concept that physicians and a machine working in combination have the greatest potential to improve clinical decision-making and patient health outcomes.

Finally, this field is still evolving and thus the entire industry may have to take a wait and see approach before settling on the right set of policies and regulations to mainstream AI in health care.

## FUNDING INFORMATION

No funding was provided for the preparation of this paper.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

## REFERENCES

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43. doi:10.1136/svn-2017-000101.
2. 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.
3. 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.
4. How Is AI Used In Healthcare - 5 Powerful Real-World Examples That Show The Latest Advances n.d. <https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/#288a84775dfb> (accessed October 14, 2018).
5. 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).
6. Healthcare's digital future | McKinsey n.d. <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future> (accessed February 3, 2019).
7. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare. Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '15, New York, New York, USA: ACM Press; 2015, p. 1721–30. doi:10.1145/2783258.2788613.
8. Bias in artificial intelligence and the importance of independent data » Brave New Coin n.d. <https://bravenewcoin.com/insights/bias-in-artificial-intelligence-and-the-importance-of-independent-data> (accessed February 12, 2019).
9. Finlayson SG, Chung HW, Kohane IS, Beam AL. Adversarial Attacks Against Medical Deep Learning Systems 2018.
10. NTSB releases preliminary report on fatal Tesla crash on Autopilot - Electrek n.d. <https://electrek.co/2018/06/07/tesla-fatal-crash-autopilot-ntsb-releases-preliminary-report/> (accessed November 1, 2018).
11. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning 2017.
12. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *Npj Digit Med* 2018;1:18. doi:10.1038/s41746-018-0029-1.
13. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* 2018;33:2358–61. doi:10.1016/J.ARTH.2018.02.067.
14. The Johns Hopkins Hospital Launches Capacity Command Center to Enhance Hospital Operations - 10/26/2016 n.d. [https://www.hopkinsmedicine.org/news/media/releases/the\\_johns\\_hopkins\\_hospital\\_launches\\_capacity\\_command\\_center\\_to\\_enhance\\_hospital\\_operations](https://www.hopkinsmedicine.org/news/media/releases/the_johns_hopkins_hospital_launches_capacity_command_center_to_enhance_hospital_operations) (accessed October 14, 2018).
15. The Dark Secret at the Heart of AI - MIT Technology Review n.d. <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> (accessed December 23, 2017).
16. 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).
17. Privacy | HHS.gov n.d. <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html> (accessed December 24, 2017).
18. 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* 2018;9:167. doi:10.3389/fgene.2018.00167.

19. Luna DNA Is About to Revolutionize Medicine -- and It'll Pay You to Help -- The Motley Fool n.d. <https://www.fool.com/investing/2018/07/31/luna-dna-is-about-to-revolutionize-medicine-and-it.aspx> (accessed October 21, 2018).
20. Artificial Intelligence, Medical Malpractice, and the End of Defensive Medicine | Bill of Health n.d. <https://blogs.harvard.edu/billofhealth/2017/01/26/artificial-intelligence-medical-malpractice-and-the-end-of-defensive-medicine/> (accessed October 14, 2018).
21. Medical Liability and Malpractice - State Laws and Legislation n.d. <http://www.ncsl.org/research/financial-services-and-commerce/medical-liability-and-malpractice.aspx> (accessed October 14, 2018).
22. Miller FH. Medical malpractice litigation: do the British have a better remedy? *Am J Law Med* 1986;11:433-63.
23. Medical Malpractice Liability: United Kingdom (England and Wales) | Law Library of Congress n.d. <https://www.loc.gov/law/help/medical-malpractice-liability/uk.php#f8> (accessed November 3, 2018).
24. 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.
25. 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/](https://www.reddit.com/r/askscience/comments/9d5id2/askscience_ama_series_im_michael_abramoff_a/) (accessed October 21, 2018).
26. EUR-Lex - 32016R0679 - EN - EUR-Lex n.d. <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679> (accessed November 3, 2018).
27. Japan pushes for basic AI rules at G-7 tech meeting | The Japan Times n.d. <https://www.japantimes.co.jp/news/2016/04/29/national/japan-pushes-basic-ai-rules-g-7-tech-meeting/#.WkXUXbaZOL4> (accessed December 28, 2017).
28. Grosz BJ, Altman R, Horvitz E, Mackworth A, Mitchell T, Mulligan D, et al. Standing Committee of the One Hundred Year Study of Artificial Intelligence n.d.
29. The National Artificial Intelligence Research and Development Strategic Plan n.d.
30. The use of algorithms in decision making: RSS evidence to the House of Commons Science and Technology Select Committee inquiry n.d.
31. ACM by U. Statement on Algorithmic Transparency and Accountability 2017.
32. ALGORITHMS AND COLLUSION Competition policy in the digital age n.d.
33. 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).
34. Focus Group on "Artificial Intelligence for Health" n.d. <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx> (accessed October 15, 2018).
35. Goodman BW. A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection n.d.
36. G7 ICT Ministers' Meeting in Takamatsu, Kagawa Joint Press Conference n.d.
37. NSTC Documents & Reports | The White House n.d. <https://obamawhitehouse.archives.gov/administration/eop/ostp/nstc/docsreports> (accessed November 3, 2018).
38. The Networking and Information Technology Research and Development (NITRD) Program n.d. <https://www.nitrd.gov/> (accessed November 3, 2018).



# 9

# CHAPTER

SUMMARY AND  
GENERAL DISCUSSION OF MAIN FINDINGS



This final chapter summarizes and discusses the findings of the work presented in this thesis.

The main aim of this thesis was to understand how technologies such as Artificial Intelligence (AI) are having an impact on healthcare. The goal was to identify challenges, knowledge gaps and propose solutions to successfully transition AI use cases from research to routine clinical care.

## SUMMARY OF MAIN FINDINGS

In order to get a good understanding of AI there is a need to explain the technology and types of AI. We then need to understand what is driving the need for AI and what are the challenges in implementing AI in healthcare. Along with introducing the thesis, in **chapter 1** we defined AI, what is driving the need for AI in healthcare, shared details on the types of AI such as Machine Learning, Neural Networks, Deep Learning, Natural Language Processing and Cognitive Computing, with healthcare relevant examples from areas such as patient care, clinical decision support, medical imaging, R&D and healthcare management. We then identified the challenges with AI around transparency (also called black box), ethics, privacy and knowledge gap as the primary drivers for transitioning AI from research to routine clinical care. Finally, in order to take a deeper look into the implementation and mainstreaming of AI in healthcare, we decided to focus on two diseases – Sepsis and Cancer. Given the physiology, ability to diagnose and treat the disease, the impact on both the patient and healthcare organization in terms of cost of care and mortality, and finally the ability to collect and analyze plethora of data, we choose these two diseases.

Driven by data from modalities such as genomics, imaging and wearables, medicine is entering the digital age (1–7). As we gain a deeper understanding of the disease biology and how diseases affect an individual, we are developing targeted therapies to personalize treatments (8,9). There is a need for technologies such as AI to sift through all this data and tailor personalized treatments. In **chapter 2** we developed a consensus paper on how we can apply AI to the emerging field of personalized healthcare especially in sepsis and cancer given their disease burden. We then presented solutions such as incorporating AI into medical education and exploring AI technologies like explainable AI to make decisions transparent, interpretable, traceable, and reproducible.

Many studies have been published on a variety of clinical applications of AI for sepsis (10–19), but there is no systematic overview of the literature around applying AI methods for prognosis, predicting mortality and treatment of sepsis. In **chapter 3** we gave an overview of this literature, identified knowledge gaps and prioritized areas 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.

## 9

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 concluded 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 risk 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.

As we looked to prioritizing areas for further research in sepsis, one area is exploring the interaction between age and benefits of early antibiotics for sepsis (20–22) and whether we can use AI techniques to identify any correlation. In **chapter 4** we applied AI to real world data from the only randomized clinical trial on this subject, the PreHospital Antibiotics Against Sepsis (PHANTASI) trial (23), to show that it is still plausible that subgroups of patients can benefit from this practice. Using real-world data, we set out to identify such subgroups that experienced 28-day mortality benefits in a prehospital setting and what were the key traits driving these benefits. We used machine learning to conduct exploratory partitioning cluster analysis to identify possible subgroups of sepsis patients. We further tested the influence of several traits within these subgroups using a logistic regression model. 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 ( $OR = 0.07$ ;  $95\%-CI = 0.01-0.79$ ;  $p = 0.03$ ). It then became clear that there is an interaction between age and benefits of early antibiotics for sepsis that had not been previously reported. 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.

Similar to most real-world data, data derived from electronic medical records for patients typically exhibit wide inter-patient variability in terms of available data elements (24) (25) (26) (27) (28). This inter-patient variability leads to missing data and can present critical challenges in developing and implementing predictive AI models to underlie clinical decision support for patient-specific disease care. In **chapter 5** we proposed a novel ensemble approach to addressing missing data that we term the “meta-model” and apply the meta-model to patient specific disease prognosis. Using real-world data from advanced lung, colorectal and breast cancer, we developed a novel machine learning based strategy to underlie clinical decision support and predict survival in cancer patients, despite the missing data. Individual models varied by the predictor data used. We combined models for each cancer type into a meta-model that predicted survival for each patient using a weighted mean of the individual models for which the patient had all requisite predictors. The meta-model significantly outperformed many of the individual models

and performed similarly to the best performing individual models. Overall, the random forest-based models performed quite well and generally outperformed the linear models in predicting survival in the held-out testing partition. The best meta-model achieved an AUC as high as 0.8 in predicting survival in test patients at key time points (500, 1000 and 1500 days). Comparisons of the meta-model to a more traditional imputation-based method of addressing missing data supported the meta-model's utility. Finally, the meta-model may address other challenges in clinical predictive modeling including model extensibility and integration of predictive algorithms trained across different institutions and datasets.

Having explored how AI is used in sepsis and cancer from a research standpoint, we sought to understand what are the challenges in transitioning AI from research to routine clinical care. Existing literature suggests that the main challenges include (29,30) - lack of accurate and sufficient data to develop and test AI models leading to increase bias in algorithms which is compounded by the lack of transparency in the models (black box effect) (31), privacy and control challenges over data (32), lack of knowledge about AI among the healthcare community (33) and the lack of policies and regulations around the use of AI from a patient care and liability standpoint (33). In the next three chapters we decided to first test these findings by conducting a survey of healthcare professionals. Second, we decided to understand the gaps in current medical education and propose a framework to introduce AI in medical education. Finally, we did a literature search of what policies and regulations are being developed and implemented to mainstream AI in routine clinical care.

In **chapter 6** we aimed to understand what AI is good for, how to evaluate it, what are its limitations, and how it can be implemented in a healthcare setting. With a survey, we aimed to evaluate the thoughts of stakeholders in laboratory medicine on the value of AI in the diagnostics space and identify anticipated challenges and solutions to introducing AI. We conducted a web-based survey on the use of AI with participants from Roche's Strategic Advisory Network that included key stakeholders in laboratory medicine. Most of the participants were medical practitioners (26%) or laboratory managers (22%). AI is currently used in the organizations of 15.6%, while 66.4% felt they might use it in the future. Most had an unsure attitude on what they would need to adopt AI in the diagnostics space. High investment costs, lack of proven clinical benefits, number of decision makers involved, and privacy concerns were identified as barriers to adoption. Education in the value of AI, streamlined implementation and integration into existing workflows, and research to prove clinical utility were identified as solutions needed to mainstream AI in laboratory medicine.

AI is not part of medical education curriculum today. Physicians go through extensive periods of training before they can eventually register as specialists. In spite of advances and changes in medicine over the past decade, traditional medical curricula has largely stayed the same (34). 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 (35). Introducing

medical students and residents to new technologies such as telemedicine, mobile healthcare applications, robotics and AI is slowly developing (34) (35) (36). In **chapter 7** we conducted a literature study of AI training in medical education, clinical curriculum and continuous medical education. We discussed the findings and proposed a training framework that included recommendations of suitable training programs along with suggested content during the various stages of medical education to incorporate AI.

Along with impressive advances and the realistic potential to transform healthcare in the near term there are difficult questions about liability and accountability, algorithmic bias and representative data, and the ability to accurately interpret and explain data that needs to be addressed to mainstream AI (37) (38,39). One key area to address is policy and regulation. In **Chapter 8** we addressed the challenges of implementing AI in routine health care practice by looking at the role of data and algorithms and the implications for policy, regulation and medical malpractice. Numerous governments, consortiums, academic and scientific groups have assembled expert stakeholders from multiple disciplines to work toward recommending action steps to alleviate these concerns. A critical component of these recommendations is making long-term investments in AI research along with understanding and addressing the ethical, legal and societal implications of AI. Additional steps include creating awareness of biases and potential harm along with explaining the procedures used by an algorithm and how a decision was made. Establishing independent councils to provide impartial advice to the government, the private sector and the public on topics related to use of algorithms will also help alleviate concerns with taking AI mainstream. For health care professionals, the recommendations by the American Medical Association (AMA) may hold the most appeal, as they advance the concept that physicians and a machine working in combination have the greatest potential to improve clinical decision-making and patient health outcomes.

## GENERAL DISCUSSION

AI took roots through Sir Alan Turing's seminal paper *Computing Machinery and Intelligence* published in journal *Mind* in 1950 , where he proposed the question, "Can machines think?" (40).

Since then AI has come a long way and today simplifies the lives of patients, doctors and hospital administrators by performing tasks that are typically done by humans, but in less time and at a fraction of the cost.

We need to understand the state of the healthcare today to truly appreciate what will drive AI. In many developed countries, mature but aged national healthcare services are being burdened with a growing aging population, accessibility to healthcare, changes in payment reforms, need for better diagnostics and treatments, worker shortage and rising costs of delivering care (41) (42) (43). Combined with a sudden surge in innovative technologies such as AI which can help with automating medical record to provide a more accurate diagnosis and tailored treatments, today's healthcare systems are ready for change (44). It is expected that the use of AI in healthcare global market is expected to grow from \$2.1 billion in 2018 to \$36.1 billion by 2025 (45).

AI is a constellation of technologies that include machine learning (46), neural networks (47), deep learning (48), natural language processing (49) and cognitive computing (50). Today AI is gaining prominence in healthcare and medicine. Numerous applications are in use in the various discipline of healthcare – clinical decision support, risk and payment management, virtual assistants, mental health, pharmaceutical and wellness. But these applications have not gone mainstream or scaled across the entire healthcare ecosystem.

## **Application of AI in Personalized Healthcare**

Medicine has entered the digital era of personalized healthcare (63) (64), driven by data from new modalities, especially genomics and imaging, as well as new sources such as wearables and Internet of Things. As we gain a deeper understanding of the disease biology and how diseases affect an individual, we are developing targeted therapies to personalize treatments. There is a need for technologies like AI to be able to support predictions for personalized treatments. In order to mainstream AI in healthcare we will need to address issues such as explainability, liability and privacy. Developing explainable algorithms and including AI training in medical education are many of the solutions that can help alleviate these concerns.

In **chapter 2** we looked at how personalized healthcare and AI are evolving. As we understand more about the biology, diagnostics, and augment medical knowledge with patient data from images, genomics, and medical records, we will be able to identify personalized therapies for individuals. Plus, with rapid digitalization of medical records (65) (66) and medical knowledge increasingly fast doubling itself, now at a rate of every 73 days (66), it is imperative that there is a need for a more robust technology to augment decision making for the various healthcare stakeholders. Based on the disease burden – financial cost, mortality, morbidity and quality-adjusted life years, we looked into two disease states namely lung cancer (67) and sepsis (68) to show what AI in combination with other technologies can or cannot do .

Following a 2018 narrative review, we found AI is widely used for the diagnosis of non-small cell lung cancer (69). Machine learning algorithms can be used to increase our understanding of important genomic pathways with the use of microarray data (70) and predict which patient will respond to newly developed checkpoint inhibitors (71) or personalize radiation therapy (72), thereby choosing an optimal treatment strategy. On the other hand, sepsis is still not well understood. A narrative review of AI applications for sepsis was published in 2019 (73), showing that applications to improve diagnosis, treatment and prognosis exist already. Many algorithms to predict sepsis onset have been developed, with encouraging results (74) (75). However, there are no specific molecular abnormalities on which new algorithms can be trained. The rapid onset and heterogeneous presentation of this syndrome makes it so, that the understanding of pathophysiology remains poor when compared to that of lung cancer. The current potential of AI is therefore limited, as unique features needed to do adequate predictions are not yet known.

We also researched the implications on patient privacy which is another outstanding issue with the use of AI. Vast amounts of patient data are needed for some AI algorithms to properly function. Google for example is using 46 billion data points collected from 216,221 adults' de-identified data over 11 combined years from two hospitals to predict the outcomes of hospitalized patients (76) (77). This raises questions about how this data is obtained and whether all patients have had a fair chance to decide about the use of their data.

Finally, we choose to look at patients as a key stakeholder in their diagnosis and treatment journey. Chatbots (78) have emerged as a key tool but physicians will need to be aware of limitations of such technologies and care for their patient accordingly (79). They will need to be trained in how to effectively use such technologies to their benefit and help ease their burden (80).

As we gain a deeper understanding of how AI works, healthcare professionals will be able to explain the decision they make with the help of AI tools. With the help of technology and regulatory bodies we will be able to resolve challenges with liability and privacy. We are well on our way to provide personalized treatment strategies driven by AI.

## 9

### Identifying Knowledge Gaps and Prioritizing Areas for Further Research – Sepsis Management Case Study

In **Chapter 3** we looked into the clinical applications of AI in sepsis by reviewing existing literature to identify knowledge gaps and prioritize areas for further research. A literature search of the electronic database PubMed was conducted and articles that reported an area under the receiver operator characteristics curve (AUROC) as an outcome measure were identified. 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.

After the initial analysis, 309 articles were extracted from PubMed and 2 additional articles were included based on expert opinion. 96 articles made the cut after reviewing the titles and abstract. After full text screening, 37 were rejected because no AUROC was reported as an outcome measure along with 18 for no link to clinical practice, 5 not about AI, 5 not about sepsis, 3 letters to the editor and 3 with no full text (only abstracts). We finally settled on 25 articles. 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.

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 (81). 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.

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 concluded 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. Also, as discussed in chapter 2, there are no specific molecular abnormalities on which new algorithms can be trained. The rapid onset and heterogeneous presentation of this syndrome makes it so, that the understanding of pathophysiology remains poor. Until these problems have been resolved, a large gap remains between the creation of an AI algorithm and its implementation in clinical practice.

### **Applying AI to Real World Data to Demonstrate Clinical Utility – Sepsis Management case study to identify subgroups of patients who may benefit from early antibiotics**

Sepsis is a major health problem worldwide. A recent study estimated the global incidence of sepsis to be nearly 50 million cases per year with 11.0 million sepsis-related deaths (82). Dysregulation of the host response to infections can cause organ dysfunction and subsequently leads to these exceptionally high mortality rates (83). Sepsis is a truly heterogeneous syndrome (84), 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.

Research has been conducted to identify specific subgroups of sepsis patients with the aim of tailoring the treatment. Seymour and colleagues, for example, categorized four clinical phenotypes of sepsis patients with similar traits, who may also respond similarly to certain treatment strategies (85). Current sepsis treatment strategies mainly focus on 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. However, it is still unclear which patients benefit most from this practice and whether antibiotics should be administered as early as possible. There is a long-standing belief that every hour of delay in administration

## 9

of antibiotics will lead to an increased risk of mortality, as suggested by Kumar et al. in 2006(86). Many treatment protocols for sepsis have been guided by this belief, ultimately resulting in an international effort called Surviving Sepsis Campaign (SSC) guideline 1-hour bundle(87) to reduce mortality and morbidity.

Recently the benefits of early antibiotic treatment in all patients with suspected sepsis have been questioned (88–90). Physicians have been forced to sacrifice diagnostic accuracy, in order to treat these patients as soon as possible, which probably contributes to overuse of antibiotics in this population (88,91,92). 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 retrospective observational studies (88). The only randomized controlled trial on this subject called Prehospital Antibiotics Against Sepsis (PHANTASi), conducted by our research group, did not show significant benefits of early administration of antibiotics in a pre-hospital setting (93).

Although there is no conclusive evidence supporting the early use of antibiotics in all patients with suspected sepsis, it might be plausible that specific subgroups of patients may benefit from early antibiotic treatment. In **chapter 4**, we aimed to identify subgroups of patients in the PHANTASi trial cohort who are likely to benefit from early antibiotic treatment. We used machine learning (73) 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.

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. A logistic regression model 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 ( $OR = 0.07$ ;  $95\%-CI = 0.007-0.75$ ;  $p = 0.03$ ).

We believe our research is novel because 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(84,94). Secondly, we have used data from the single randomized trial on this subject, which makes the chance of residual confounding as low as it could ever be for this type of study. Lastly, we had the opportunity to evaluate the effect of potential confounders such as antibiotic sensitivities, while most studies on this subject lack this very important data to evaluate adequacy of the given treatment (95).

We recognize the inherent limitations of performing secondary analyses on a data set. Subgroup effects can often be misleading and can be explained by chance(96). To minimize the risk that we found these results by chance, we performed several slightly

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.

The findings in our study may be of substantial benefit to any randomized or observational study on this subject, as researchers can now be aware that there is a potential interaction between the benefits of early antibiotic treatment and the age of the patient. We are interested to see whether adjustments for this interaction will influence the results of other studies on this subject that have already been published. Furthermore, the results of this analysis may provide new insights into the sepsis immunopathology and the role of the ageing immune system in sepsis. Insights in the altered host response during ageing may advance fundamental, as well as clinical research on sepsis immunopathology.

Along with sepsis, we choose to dive deeper into another disease state namely cancer to explore how AI can be used for diagnosis.

9

### **Applying AI to Real World Data to Demonstrate Clinical Utility – Oncology case study to address the challenge of adequate data needed to train and implement robust AI models**

With cancer, often the key to successful treatment is catching it early. In many cases, patients with clinical symptoms or findings concerning for cancer will first be evaluated using various medical imaging modalities such as CT, MRI or ultrasound. While skilled radiologists can often identify imaging findings suggestive or diagnostic of cancer and can sometimes exclude cancer, imaging alone is usually insufficient for a specific diagnosis. For example, depending on context, imaging is sometimes unable to distinguish benign from malignant lesions and usually cannot definitively identify the histologic tumor type. For definitive diagnosis, a biopsy is often required. However, even biopsies can produce false-negative or false-positive results.

Researchers and oncologists are keen to improve the diagnostic process to avoid these issues. Detecting whether a lesion is malignant or benign more reliably and without the need for a biopsy would be a game changer. This is where AI comes into the picture. Team of oncologists, researchers and computer sciences experts develop and train an algorithm to scan the various cancer registries for cancer patient diagnosis and look for certain patterns for example genetic mutations – EGFR, ALK. Cancer patients' electronic medical records, including tumor pathology reports, are customarily fed into cancer registries that go on to form the National Cancer Institute's Surveillance, Epidemiology and End Results Program, known as SEER. The SEER database (97), which serves as a proxy for the U.S. cancer patient population, provides valuable information to researchers on

## 9

cancer incidence and mortality, as well as patient demographics, tumor characteristics and treatment. Similarly in Europe, the EUROCARE (98) study monitors the survival of cancer patients in Europe through centralized collection, quality check and statistical analysis of population-based cancer registries data.

Many commonly used machine learning algorithms require complete data sets and cannot directly use for training or prediction datasets containing missing elements of predictor data. Several strategies are often employed to use real world data that include missing elements in predictive analyses. A commonly used strategy involves pre-processing a set of clinical data by “imputing” missing data elements (99). While there are numerous variations on imputation and related approaches including single imputation, multiple imputation, and expectation maximization, most imputation approaches are fundamentally designed to use available data to estimate the distribution or value of each element of missing data (100) (24,27,101) (102–105). In **Chapter 5** we aimed to develop and demonstrate a machine-learning method to calculate patient-specific cancer survival probability functions, suitable for generation of patient-specific Kaplan-Meier curves, in the setting of substantial intra-patient data availability and “missing data”. Cancer patients exhibit wide inter-patient variation with regard to which diagnostic studies and tests they have performed, limiting many traditional approaches to clinical forecasting. We developed a method for cancer prognostication applicable to patients with varying predictor data availability.

The primary objective of our research was to develop and demonstrate a novel “meta-model” approach to addressing missing data. As described in the paper, our meta-model concept includes an ensemble of underlying models based on varying predictors with the final output based on an aggregate of all individual models for which a patient has complete data. The meta-model may also address other challenges in predictive clinical decision support (CDS) implementation including model extensibility and integration of predictive algorithms trained across different institutions and datasets. The secondary objective was to develop a method for generating patient-specific Kaplan-Meier survival curves.

Using clinical data from patients with metastatic colorectal cancer, metastatic breast cancer and advanced lung cancer, we first developed a set of survival prediction models intended to individualize patient prognosis. After developing and validating the individual models, we combined the individual models for each dataset into a “meta-model.” We demonstrated that this meta-model can provide an alternative to imputation in addressing missing data and may offer several key advantages.

Using data from Flatiron Health, we developed a suite of individual linear and random survival forest models to predict survival in patients with advanced lung, colorectal (CRC) and breast cancer. Individual varied by the predictor data used. We combined individual CRC models into a “meta-model” that predicted survival for each patient using a weighted-mean of the individual models for which the patient had all requisite predictors.

We evaluated our models by comparing predicted survival to actual survival for patients in the test partition. The best performing individual models predicted survival on held out test data at 500, 1000 and 1500 days from advanced diagnosis with an AUC of >0.75 for CRC, >0.8 for breast cancer and >0.85 for lung cancer. We demonstrated the application of the meta-model in generating patient-specific survival curves.

The “meta-model” approach we developed and demonstrated in paper manuscript offers a strategy to develop clinical predictive models that can accommodate inter-patient heterogeneity in data availability and “missing data”. We further demonstrated the value of random forest-based survival models in predicting patient-specific oncology survival. We expect that the proofs of concept developed here will provide a foundation for novel types of clinical decision support to enable clinicians to make more personalized patient-care decisions.

Having understood the uses of AI in healthcare by focusing on sepsis and cancer, the next activity was to get a pulse of how AI is used today in healthcare. We decided to focus on diagnostics and laboratory medicine. Advances in our understanding of biology, disease and molecular medicine have created a central role for laboratory medicine in the diagnostic work-up of many, if not most diseases. It is estimated that in developed countries 70% of decisions regarding a patient’s diagnosis, treatment, and discharge are based on results of laboratory tests (106). Unfortunately, medical errors due to misdiagnosis is the third leading cause of deaths in the United States(U.S.) (107) (108). Another study conducted across the U.S. and Europe showed that one in 20 patients (6%) is impacted by preventable medical errors (109). The ever-increasing workload, high healthcare costs and need for improved precision, call for continuous optimization of the laboratory processes (110).

### **Identifying challenges with transitioning AI to routine clinical care – Laboratory Medicine Case Study**

As laboratory medicine continues to undergo digitalization and automation, clinical laboratorians will likely be confronted with the challenges associated with evaluating, implementing, and validating AI algorithms, both inside and outside their laboratories. Understanding what AI is good for, where it can be applied, along with the state-of-the-art and limitations will be useful to practicing laboratory professionals and clinicians. On the other hand, the introduction of new technologies requires willingness to change the current structure and mindset towards these technologies, which are not always well understood. Historically, there has been resistance to the adoption of new technologies in the medical community since the medical professionals believe that these would interfere with their ability to make independent diagnosis and their relationships with patients (111).

We conducted a survey of the laboratory medicine community which includes practicing physicians and surgeons, point of care coordinators, anatomic pathologists, laboratory management, information technology management and senior leadership.

**Chapter 6** aimed to evaluate their thoughts on the value of AI in the diagnostics space and identify anticipated challenges and potential solutions to introducing AI in this field.

The survey was conducted in a 3-step process. Firstly, 98 stakeholders participated in a two-week online discussion board on AI in diagnostics in August 2019. These participants were part of Roche's Strategic Advisory Network and included laboratory medicine decision makers, practicing physicians and surgeons, point of care coordinators, anatomic pathologists, laboratory management, information technology management and senior leadership. Next, two one-hour online group chats were organized on October 2<sup>nd</sup> and 3<sup>rd</sup> 2019 to discuss these questions and fine-tune their phrasing and to refine the answer possibilities to the multiple-choice questions. In these group chats, a total of 11 practitioners in laboratory medicine were asked to answer the initial survey questions one at the time, after which they could comment on each other's answers and discuss their opinions on and interpretations of the questions.

Finally, this thoroughly discussed survey was fielded to a group of 302 laboratory medicine practitioners via email. The survey was available from October 21<sup>st</sup> until November 1<sup>st</sup> 2019. The data was collected using a software platform called Confirmit (Oslo, Norway) and all participants gave informed consent for their input to be used for research purposes.

It became clear AI is still at its infancy in labs, used by only 15.6% of the participants today. The positive being 66.4% felt that AI might be used in the future. The respondents used AI for diagnosing diseases from images (30%), reviewing patients' risk profiles for certain conditions (40%), pre-empt rapid response solutions (30%), and for automatically releasing laboratory results and financial analytics (10%).

81% felt AI will be valuable in their organizations in the next 5 years. Interestingly the 19% who did not see value of AI to their organizations had a consistent response – very expensive, lack of capital dollars to invest in new technologies, and as simple as – I'm not sure about the use of AI.

High investment costs, lack of proven clinical benefits, large number of laboratory and clinical decision makers and patient privacy concerns were identified as top barriers to implementation and wide spread adoption. Another concern was with the number of decision makers (up to 10) involved that would decide whether an organization should pursue AI or not.

Education and training in the value of AI, streamlined implementation and integration of AI into existing workflows and research to prove out clinical utility were identified as solutions needed to mainstream AI in laboratory medicine.

The quantitative results of our survey are similar to those found in surveys on this subject amongst other healthcare professionals. Similar to our findings, surveys on AI for pathologists (112), medical students (113), physicians (114) and radiologists (115) all found that about 80% of participants feel like AI will be influential or valuable in their practice in the upcoming years. Interestingly enough, a 2019 survey shows that 84% of the general population in the US feels that AI will be at the centre of the next technological revolution (116).

The respondents also felt that AI could help with reducing costs and improving diagnostics efficiency, by say avoiding ordering the same test twice. Recent study has shown that AI can help reduce the waste in the US healthcare system in the range from \$760 billion to \$935 billion in 2019 (117). On the subject of impact on jobs - the general consensus was AI would create efficiencies that expedite their workflows, but want to ensure that they are still in control.

We also addressed the implication on patients. One of the respondents posed an interesting question in the online group chat: "*Should the patients be informed that some of the decisions are being recommended by AI?*" Another question is whether we should inform patients when an AI recommendation is not followed. Unfortunately, this burdens the patient as they now have to choose between the physician and the computer. Many algorithms are already being used in medicine, like the YEARS criteria (118) for pulmonary embolisms. Their role in the diagnostic process is rarely explained to or discussed with the patient, as only trained physicians can interpret the results of these algorithms. We therefore believe that a similar approach might be best when using more advanced algorithms, in which explainability and interpretation are an even larger problem. Finally, these tools and algorithms are an aid to complement the healthcare practitioners who are eventually responsible for the diagnostic process and decision making.

### **Addressing challenges with transitioning AI to routine clinical care – framework to augment existing medical education curriculum to include AI**

Since education of healthcare stakeholders was brought up in the survey results, we decided to understand the medical education landscape today and what it would take to introduce AI in routine clinical practice. With education being a key enabler for wide spread AI adoption, we conducted a literature study of medical education curriculum in medical colleges starting with the Medical College Admission Tests, United States Medical Licensing Examinations, Accreditation Council for Graduate Medical Education, and Continuing Medical Education (CME). **Chapter 7** discussed the findings which were very underwhelming. As the practice of medicine enters the age of 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.

We summarized a list of AI initiatives in medical education and training available in AI with CME credits. Unfortunately, this list isn't long.

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 (36). To embed this knowledge into medicine, it is of the essence to start introducing these concepts from the beginning of training. In many countries, 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 (119). 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 (120), provenance (121), integration (122) and governance, working with EHR's (123), AI fundamentals, and ethics and legal issues with AI (124) (125). An approach to introducing AI could be to incorporate this technology during courses such as Evidence Based Medicine (126).

During clinical rotations and residency, focus should shift more towards relevant applications of AI in practice. With advancements in digital biomarkers (127) and digital therapeutics (128), 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 (129). 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.

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 (130). 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 (131). These projects are new developments and are the first steps taken in order to introduce AI in medical education.

It is clear that 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 (36). Less time is spent on familiarizing medical students or residents with new technologies such as AI, mobile health care applications and telemedicine (34) (35) (36). 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 (132).

## Addressing challenges with transitioning AI to routine clinical care – policy and regulatory changes needed to enable safe use of AI

We then decided to start looking into what are the policy and regulatory implications of mainstreaming AI in healthcare. Computer algorithms are at the core of AI. These algorithms are being used to make health care decisions such as prioritizing activities for staff members or triggering interventions for admitted patients—as was reported at John Hopkins Hospital (133). In these situations, how can one trust the algorithm to do the right thing? Take the case of ‘Deep Patient’ where researchers at Mount Sinai Hospital applied deep learning to 700,000 patient records. Without expert guidance the tool was able to identify patterns and predict the onset of diseases such as cancer of the liver. The tool also predicted the onset of schizophrenia but offered no clue as to how it did so. For a condition that is very difficult to predict even by the most experienced psychiatrist, the way the AI system came up with its decision is known as the “black box” problem (134). So how can we trust such a system?

Even more concerning, what are the consequences for the patient? If physicians tell their patients they are going to develop schizophrenia in the future and over time they do not develop it, what impacts does this have on the patients? This raises another ethical dilemma as well: when physicians realize they made a wrong diagnosis due to AI, do they have an obligation to tell their patients about it?

Decisions made by predictive algorithms can be obscure because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations).

In considering the application of algorithms for individual patient treatment decisions, it should be noted that the physician will have to interpret the prediction model, as it is created by running the same algorithm on a mass scale. The physician will have to decide that the prediction score is reliable and accordingly propose a diagnosis.

The AI machine will be able to predict diagnosis based on complex relationships between the patient and expected treatment results without explicitly identifying or understanding those connections. We are now entering an era where the medical decision-making burden shifts from the physician to an algorithm. What happens when the physician pursues an improper treatment—based on an algorithm—that results in an error? Who is liable for erroneous care based on a decision made by a machine is a hard problem to solve? Future malpractice guidelines will need to incorporate such considerations by including health care professionals but also software companies that created the algorithm.

In **chapter 8** we thoughtfully examined these issues and challenges and found out that as AI evolves rapidly, the adoption and diversity of applications in medical practice are outpacing the critical need of establishing standards and guidelines to protect the health care community. There is a need for key stakeholders from both the public and private sector to collaborate and recommend policy guidelines to enable the safe use of AI. We

## 9

compiled a list of 9 global initiatives focused on developing AI policy guidelines. A few activities include -

The Global Data Protection Regulation (135) from the European Union explicitly addresses algorithmic discrimination by – “Data Sanitization” – removing special categories from data sets used in automated decision making and “Right to Explanation” whereby data subjects are entitled to “meaningful information about the logic involved, as well as the significance and the envisaged consequences” when automated decision making or profiling takes place. This has now been enacted into legislation. The National Science and Technology Council (136), also published in 2016, created a subcommittee on machine learning and AI, and drafted a strategy to coordinate all US government activities in AI. A 7-point action plan with recommendations especially for healthcare included provisions for transparency and explainability. The Royal Statistical Society (137) in the UK published its findings in 2017 and recommended setting up an independent data ethics council to provide advice to government, public and private sector on the use of algorithms. Finally, in 2018 the American Medical Association (138) published a report saying that the overarching goal of AI in health care is to be human-centered and augment human intelligence and advance the quadruple aim: improve population health; improve health outcomes and patient satisfaction; increase value; and improve health care team satisfaction.

To summarize our work, numerous governments, consortiums and academic or scientific groups have assembled expert stakeholders from multiple disciplines to work toward recommending action steps to alleviate these concerns. A critical component of these recommendations is making long-term investments in AI research along with understanding and addressing the ethical, legal and societal implications of AI. Additional steps include creating awareness of biases and potential harm along with explaining the procedures used by an algorithm and how a decision was made. Establishing independent councils to provide impartial advice to the government, the private sector and the public on topics related to use of algorithms will also help alleviate concerns with taking AI mainstream. In the end, the barrier to wider AI adoption should not be the notion that AI has to be perfect (right ‘every time’). We should focus on demonstrating that a doctor with AI is better than a doctor without AI and not the AI has to be perfect.

## REFERENCES

1. Adamo JE, Bienvenu li R V, Dolz F, Liebman M, Nilsen W, Steele SJ. Translation of Digital Health Technologies to Advance Precision Medicine: Informing Regulatory Science. *Digit biomarkers [Internet]*. 2020 Feb 7 [cited 2021 Apr 26];4(1):1–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32399511>
2. Chen YW, Nishikawa I, Tamura S, Lu BL, Jiang H. Computational intelligence in biomedical science and engineering. Vol. 2012, *Computational Intelligence and Neuroscience*. 2012.
3. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Vol. 286, *Radiology*. Radiological Society of North America Inc.; 2018. p. 800–9.
4. Sacha GM, Varona P. Artificial intelligence in nanotechnology. Vol. 24, *Nanotechnology*. 2013.
5. Mainali G. Artificial intelligence in medical science: perspective from a medical student. Vol. 58, *Journal of the Nepal Medical Association*. Nepal Medical Association; 2020. p. 709–71.
6. Kritz F. Technology is transforming work for nurses and care for patients. Vol. 119, *American Journal of Nursing*. Lippincott Williams and Wilkins; 2019. p. 18–9.
7. Medicine in the digital age [Internet]. Vol. 25, *Nature Medicine*. Nature Publishing Group; 2019 [cited 2021 Apr 26]. Available from: <https://pubmed.ncbi.nlm.nih.gov/30617338/>
8. Krzyszczuk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized medicine for cancer treatment. *TECHNOLOGY [Internet]*. 2018 Sep [cited 2021 Apr 26];06(03n04):79–100. Available from: <https://pubmed.ncbi.nlm.nih.gov/30713991/>
9. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress [Internet]. Vol. 109, *Fertility and Sterility*. Elsevier Inc.; 2018 [cited 2021 Apr 26]. p. 952–63. Available from: [/pmc/articles/PMC6366451/](https://pmc/articles/PMC6366451/)
10. Sun Y, Kaur R, Gupta S, Paul R, Das R, Cho SJ, et al. Development and validation of high definition phenotype-based mortality prediction in critical care units. *JAMIA open [Internet]*. 2021 Jan 1 [cited 2021 Apr 26];4(1):ooab004. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33796821>
11. 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 [Internet]*. 2018 Apr 1 [cited 2021 Apr 26];46(4):547–53. Available from: <https://pubmed.ncbi.nlm.nih.gov/29286945/>
12. Shashikumar SP, Josef CS, Sharma A, Nemati S. DeepAISE – An interpretable and recurrent neural survival model for early prediction of sepsis. *Artif Intell Med*. 2021 Mar 1;113.
13. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, 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 Jan 1;8(1).
14. Johnson AEW, Aboab J, Raffa JD, Pollard TJ, Deliberato RO, Celi LA, et al. A Comparative Analysis of Sepsis Identification Methods in an Electronic Database. *Crit Care Med*. 2018 Apr 1;46(4):494–9.
15. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Vol. 46, *Intensive Care Medicine*. Springer; 2020. p. 383–400.
16. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc*. 2018 Jul 1;15(7):846–53.
17. Wong A-KI, Cheung PC, Kamaleswaran R, Martin GS, Holder AL. Machine Learning Methods to Predict Acute Respiratory Failure and Acute Respiratory Distress Syndrome. *Front big data [Internet]*. 2020 Nov 23 [cited 2021 Apr 26];3:579774. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33693419>

18. Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. Vol. 33, Journal of Clinical Monitoring and Computing. Springer Netherlands; 2019. p. 887–93.
19. Ding M, Luo Y. Unsupervised phenotyping of sepsis using nonnegative matrix factorization of temporal trends from a multivariate panel of physiological measurements. BMC Med Inform Decis Mak. 2021 Apr;21(S5).
20. Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. N Engl J Med. 2017 Jun 8;376(23):2223–34.
21. Singer M. Antibiotics for sepsis: Does each hour really count, or is it incestuous amplification? [Internet]. Vol. 196, American Journal of Respiratory and Critical Care Medicine. American Thoracic Society; 2017 [cited 2021 Apr 26]. p. 800–2. Available from: [www.atsjournals.org](http://www.atsjournals.org).
22. 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 Oct 1 [cited 2021 Apr 26];196(7):856–63. Available from: [/pmc/articles/PMC5649973/](http://pmc/articles/PMC5649973/)
23. 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 1 [cited 2020 Aug 29];6(1):40–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29196046>
24. Weber GM, Adams VG, Bernstam E V, Bickel JP, Fox KP, Marsolo K, et al. Biases introduced by filtering electronic health records for patients with "complete data". J Am Med Inform Assoc [Internet]. 2017 Nov 1 [cited 2020 Nov 28];24(6):1134–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29016972>
25. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open [Internet]. 2013 Aug 1 [cited 2020 Mar 8];3(8):e002847. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23906948>
26. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. Am J Clin Pathol [Internet]. 2016 Jun [cited 2019 Aug 9];145(6):778–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27329638>
27. Luo Y, Szolovits P, Dighe AS, Baron JM. 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. J Am Med Inform Assoc [Internet]. 2018 [cited 2020 Nov 28];25(6):645–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29202205>
28. Baron JM, Kurant DE, Dighe AS. Machine Learning and Other Emerging Decision Support Tools. Clin Lab Med [Internet]. 2019 Jun [cited 2020 Mar 8];39(2):319–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31036284>
29. Park CW, Seo SW, Kang N, Ko BS, Choi BW, Park CM, et al. Artificial Intelligence in Health Care: Current Applications and Issues [Internet]. Vol. 35, World Journal of Orthopedics. Baishideng Publishing Group Co; 2020 [cited 2021 Apr 26]. p. 1–11. Available from: [/pmc/articles/PMC7606883/](http://pmc/articles/PMC7606883/)
30. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence [Internet]. Vol. 17, BMC Medicine. BioMed Central Ltd.; 2019 [cited 2021 Apr 26]. Available from: [/pmc/articles/PMC6821018/](http://pmc/articles/PMC6821018/)
31. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. PLoS One [Internet]. 2020 [cited 2021 Feb 16];15(4):e0231166. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32251471>
32. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. Front Public Heal [Internet]. 2016 Feb 17 [cited 2018 Sep 15];4:7. Available from: <http://journal.frontiersin.org/Article/10.3389/fpubh.2016.00007/abstract>
33. Hazarika I. Artificial intelligence: Opportunities and implications for the health workforce. Int Health [Internet]. 2020 [cited 2021 Apr 26];12(4):241–5. Available from: [/pmc/articles/PMC7322190/](http://pmc/articles/PMC7322190/)

34. Pfeifer CM. A progressive three-phase innovation to medical education in the United States. *Med Educ Online*. 2018;00.
35. Nasca TJ, Philibert I, Brigham T, Flynn TC. The Next GME Accreditation System — Rationale and Benefits. *N Engl J Med*. 2012 Mar;366(11):1051–6.
36. Wartman SA, Donald Combs C. Medical education must move from the information age to the age of artificial intelligence. *Acad Med [Internet]*. 2018 [cited 2021 Apr 26];93(8):1107–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/29095704/>
37. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: *Artificial Intelligence in Healthcare [Internet]*. Elsevier; 2020 [cited 2021 Apr 26]. p. 295–336. Available from: [/pmc/articles/PMC7332220/](https://pmc/articles/PMC7332220/)
38. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: Accountability and safety. *Bull World Health Organ [Internet]*. 2020 Apr 1 [cited 2021 Apr 26];98(4):251–6. Available from: [/pmc/articles/PMC7133468/](https://pmc/articles/PMC7133468/)
39. Smith H. Clinical AI: opacity, accountability, responsibility and liability. *AI Soc [Internet]*. 2020 Jul 25 [cited 2021 Apr 26];1:3. Available from: <https://doi.org/10.1007/s00146-020-01019-6>
40. Alan M. Turing. Computing machinery and intelligence. *Mind*. 1950;
41. Vogenberg FR, Santilli J. Healthcare Trends for 2018. *Am Heal drug benefits [Internet]*. 2018 Feb [cited 2021 Feb 15];11(1):48–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29692880>
42. 2020 Global health care sector outlook | Deloitte [Internet]. [cited 2021 Feb 15]. Available from: <https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html>
43. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: Systematic review [Internet]. Vol. 22, *Journal of Medical Internet Research*. JMIR Publications Inc.; 2020 [cited 2021 Apr 22]. Available from: [/pmc/articles/PMC7059082/](https://pmc/articles/PMC7059082/)
44. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Futur Healthc J [Internet]*. 2019 Jun [cited 2021 Feb 15];6(2):94–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31363513>
45. Artificial Intelligence in Healthcare Spending to Hit \$36B [Internet]. [cited 2021 Apr 22]. Available from: <https://healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b>
46. Louridas P, Ebert C. Machine Learning. *IEEE Softw [Internet]*. 2016 Sep [cited 2020 Jan 30];33(5):110–5. Available from: <https://ieeexplore.ieee.org/document/7548905/>
47. Neural Networks in Healthcare [Internet]. [cited 2017 Jul 28]. Available from: <http://www.openclinical.org/neuralnetworksinhealthcare.html>
48. Du X, Cai Y, Wang S, Zhang L. Overview of deep learning. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC) [Internet]. IEEE; 2016 [cited 2020 Jan 30]. p. 159–64. Available from: <https://ieeexplore.ieee.org/document/7804882/>
49. Kłosowski P. Deep Learning for Natural Language Processing and Language Modelling. In: 2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA) [Internet]. IEEE; 2018 [cited 2020 Jan 30]. p. 223–8. Available from: <https://ieeexplore.ieee.org/document/8563389/>
50. Chen Y, Argentinis E, Weber G. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research [Internet]. Vol. 38, *Clinical Therapeutics*. Excerpta Medica Inc.; 2016 [cited 2021 May 21]. p. 688–701. Available from: <https://pubmed.ncbi.nlm.nih.gov/27130797/>
51. Ta CN, Weng C. Detecting systemic data quality issues in electronic health records. In: *Studies in Health Technology and Informatics [Internet]*. IOS Press; 2019 [cited 2021 May 13]. p. 383–7. Available from: [/pmc/articles/PMC6857180/](https://pmc/articles/PMC6857180/)
52. Stanfill MH, Marc DT. Health Information Management: Implications of Artificial Intelligence on Healthcare Data and Information Management [Internet]. Vol. 28, *Yearbook of medical informatics*. NLM (Medline); 2019 [cited 2021 May 13]. p. 56–64. Available from: [/pmc/articles/PMC6697524/](https://pmc/articles/PMC6697524/)

53. Finlayson SG, Chung HW, Kohane IS, Beam AL. Adversarial Attacks Against Medical Deep Learning Systems. 2018 Apr 14 [cited 2019 Feb 3]; Available from: <http://arxiv.org/abs/1804.05296>
54. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Eng* [Internet]. 2017 Feb 10 [cited 2021 May 21];1(2):1–8. Available from: <https://www.nature.com/articles/s41551-016-0024>
55. 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* [Internet]. 2018 Dec 28 [cited 2021 May 21];1(1):39. Available from: [www.nature.com/npjdigitalmed](http://www.nature.com/npjdigitalmed)
56. Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* [Internet]. 2018 Oct 1 [cited 2021 May 21];2(10):741–8. Available from: <https://www.nature.com/articles/s41551-018-0301-3>
57. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *Am Heart J* [Internet]. 2019 Jan 1 [cited 2021 May 21];207:66–75. Available from: <https://pubmed.ncbi.nlm.nih.gov/30392584/>
58. Willyard C. Can AI Fix Medical Records? Vol. 576, *Nature*. Nature Research; 2019. p. S59–62.
59. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. Brock G, editor. *PLoS One* [Internet]. 2015 Mar 4 [cited 2021 Apr 30];10(3):e0118432. Available from: <https://dx.plos.org/10.1371/journal.pone.0118432>
60. Price WN. Regulating Black-Box Medicine. *Mich Law Rev* [Internet]. [cited 2020 Jan 30];116(3):421–74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29240330>
61. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* [Internet]. 2019 [cited 2020 Jan 30];28(3):231–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30636200>
62. Ethics and Governance of AI at Berkman Klein: Report on Impact, 2017-2019 | Berkman Klein Center [Internet]. [cited 2020 Jan 30]. Available from: <https://cyber.harvard.edu/story/2019-10/ethics-and-governance-ai-berkman-klein-report-impact-2017-2019>
63. Agyeman AA, Ofori-Asenso R. Perspective: Does personalized medicine hold the future for medicine? *J Pharm Bioallied Sci* [Internet]. 2015 [cited 2020 Jan 30];7(3):239–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26229361>
64. Mathur S, Sutton J. Personalized medicine could transform healthcare. *Biomed reports* [Internet]. 2017 Jul [cited 2020 Jan 30];7(1):3–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28685051>
65. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Aff* [Internet]. 2015 Dec 2 [cited 2019 Jul 27];34(12):2174–80. Available from: <http://www.healthaffairs.org/doi/10.1377/hlthaff.2015.0992>
66. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015 [Internet]. [cited 2019 Nov 25]. Available from: <https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php>
67. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* [Internet]. 2018 Nov 1 [cited 2020 Nov 26];68(6):394–424. Available from: <http://doi.wiley.com/10.3322/caac.21492>
68. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study.

- Lancet (London, England) [Internet]. 2020 Jan 18 [cited 2020 Nov 26];395(10219):200–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31954465>
69. Rabbani M, Kanevsky J, Kafi K, Chandelier F, Giles FJ. Role of artificial intelligence in the care of patients with nonsmall cell lung cancer. Vol. 48, European Journal of Clinical Investigation. Blackwell Publishing Ltd; 2018.
  70. Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, et al. Expression profiles of non-small cell lung cancers on cDNA microarrays: Identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogene. 2003 Apr;22(14):2192–205.
  71. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017 Jan;18(1):248–62.
  72. Amit G, Purdie TG, Levishtein A, Hope AJ, Lindsay P, Marshall A, et al. Automatic learning-based beam angle selection for thoracic IMRT. Med Phys. 2015 Apr;42(4).
  73. Schinkel M, Paranjape K, Panday RSN, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med [Internet]. 2019 Oct 7 [cited 2019 Oct 16];103488. Available from: <https://www.sciencedirect.com/science/article/pii/S0010482519303567>
  74. 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;
  75. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, 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.
  76. Google is using 46 billion data points to predict a hospital patient's future — Quartz [Internet]. [cited 2018 Nov 1]. Available from: <https://qz.com/1189730/google-is-using-46-billion-data-points-to-predict-the-medical-outcomes-of-hospital-patients/>
  77. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Liu PJ, et al. Scalable and accurate deep learning for electronic health records. 2018 Jan 24 [cited 2019 Aug 26]; Available from: <http://arxiv.org/abs/1801.07860>
  78. Bates M. Health Care Chatbots Are Here to Help. IEEE Pulse [Internet]. 2019 May [cited 2019 Aug 10];10(3):12–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31135345>
  79. 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 [Internet]. 2019 Apr 5 [cited 2019 Aug 10];21(4):e12887. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30950796>
  80. How Chatbots Can Ease Primary Care Burden | Medpage Today [Internet]. [cited 2019 Aug 10]. Available from: <https://www.medpagetoday.com/primarycare/generalprimarycare/72802>
  81. Moons KGM, Wolff RF, Riley RD, Penny ;, Whiting F, Westwood M, 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.
  82. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. Lancet (London, England) [Internet]. 2020 Jan 18 [cited 2020 Feb 11];395(10219):200–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31954465>
  83. 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. 2016 Feb;315(8):801.
  84. de Groot 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. Vol. 44, Intensive Care Medicine. Springer Verlag; 2018. p. 311–22.

85. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019 May;321(20):2003.
86. 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*. 2006 Jun;34(6):1589–96.
87. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med*. 2017;43:304–77.
88. 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*. 2019;
89. 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.
90. 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*. 2019 Apr;73(4):356–8.
91. 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 Jun;75(5):196–203.
92. 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;19(1).
93. 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 2020 Feb 10];6(1):40–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29196046>
94. 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 Sep;43(9):1907–15.
95. 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.
96. Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis users' guides to the medical literature. Vol. 311, JAMA - Journal of the American Medical Association. American Medical Association; 2014. p. 405–11.
97. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin [Internet]*. 2020 Jan 8 [cited 2020 Mar 14];70(1):7–30. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21590>
98. Rossi S, Baili P, Capocaccia R, Caldora M, Carrani E, Minicozzi P, et al. The EUROCare-5 study on cancer survival in Europe 1999-2007: Database, quality checks and statistical analysis methods. *Eur J Cancer [Internet]*. 2015 Oct [cited 2020 Nov 26];51(15):2104–19. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26421815>
99. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and Managing Missing Structured Data in Electronic Health Records: Data Analysis. *JMIR Med informatics [Internet]*. 2018 Feb 23 [cited 2020 Nov 28];6(1):e11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29475824>
100. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. *Am J Clin Pathol [Internet]*. 2016 Jun [cited 2020 Nov 28];145(6):778–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27329638>

101. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* [Internet]. 2013 Aug 1 [cited 2020 Nov 28];3(8):e002847. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23906948>
102. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* [Internet]. 2011 Mar [cited 2020 Nov 28];20(1):40–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21499542>
103. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* [Internet]. 1999 Mar 30 [cited 2020 Nov 28];18(6):681–94. Available from: [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0258\(19990330\)18:6%3C681::AID-SIM71%3E3.0.CO;2-R](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(19990330)18:6%3C681::AID-SIM71%3E3.0.CO;2-R)
104. Qi L, Wang Y-F, He Y. A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Stat Med* [Internet]. 2010 Nov 10 [cited 2020 Nov 28];29(25):2592–604. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20806403>
105. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* [Internet]. 2007 Feb [cited 2020 Nov 28];61(1):79–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17401454>
106. Badrick T. Evidence-based laboratory medicine. *Clin Biochem Rev* [Internet]. 2013 Aug [cited 2019 Dec 22];34(2):43–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24151340>
107. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* [Internet]. 2016 May 3 [cited 2019 Dec 30];353. Available from: <https://www.bmjjournals.org/content/353/bmjjournals.12139.long>
108. To Err Is Human [Internet]. Washington, D.C.: National Academies Press; 2000 [cited 2019 Dec 30]. Available from: <http://www.nap.edu/catalog/9728>
109. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ* [Internet]. 2019 Jul 17 [cited 2020 Nov 26];366:l4185. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31315828>
110. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. *Crit Rev Clin Lab Sci* [Internet]. 2019 Feb 17 [cited 2019 Aug 9];56(2):98–110. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30922144>
111. Safi S, Thiessen T, Schmailzl KJ. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR Res Protoc*. 2018 Dec;7(12):e11072.
112. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *npj Digit Med* [Internet]. 2019 Dec 26 [cited 2019 Aug 9];2(1):28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31304375>
113. Pinto dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. *Eur Radiol* [Internet]. 2019 Apr 6 [cited 2019 Dec 22];29(4):1640–6. Available from: <http://link.springer.com/10.1007/s00330-018-5601-1>
114. Oh S, Kim JH, Choi S-W, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. *J Med Internet Res* [Internet]. 2019 [cited 2019 Dec 22];21(3):e12422. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30907742>
115. Collado-Mesa F, Alvarez E, Arheart K. The Role of Artificial Intelligence in Diagnostic Radiology: A Survey at a Single Radiology Residency Training Program. *J Am Coll Radiol* [Internet]. 2018 Dec [cited 2019 Dec 22];15(12):1753–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1546144017316666>
116. 2019 Artificial Intelligence Survey | Edelman [Internet]. [cited 2019 Dec 22]. Available from: <https://www.edelman.com/research/2019-artificial-intelligence-survey>
117. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System. *JAMA* [Internet]. 2019 Oct 15 [cited 2019 Dec 23];322(15):1501. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2752664>

118. van der Hulle T, Cheung WY, Kooij S, Beenen LFM, van Bemmel T, van Es J, et al. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. *Lancet.* 2017 Jul;390(10091):289–97.
119. The MCAT ® Essentials for Testing Year 2019.
120. van Ooijen PMA. Quality and Curation of Medical Images and Data. In: Artificial Intelligence in Medical Imaging [Internet]. Cham: Springer International Publishing; 2019 [cited 2019 Aug 8]. p. 247–55. Available from: [http://link.springer.com/10.1007/978-3-319-94878-2\\_17](http://link.springer.com/10.1007/978-3-319-94878-2_17)
121. Council Post: Four Reasons Data Provenance Is Vital For Analytics And AI [Internet]. [cited 2019 Aug 9]. Available from: <https://www.forbes.com/sites/forbestechcouncil/2019/05/22/four-reasons-data-provenance-is-vital-for-analytics-and-ai/#503e0cdb57d6>
122. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). *Methods Inf Med* [Internet]. 2018 [cited 2019 Aug 9];57(S 01):e57–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30016812>
123. Wald HS, George P, Reis SP, Taylor JS. Electronic Health Record Training in Undergraduate Medical Education. *Acad Med* [Internet]. 2014 Mar [cited 2019 Aug 9];89(3):380–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24448045>
124. Price WN. Regulating Black-Box Medicine [Internet]. 2017 [cited 2019 Aug 9]. Available from: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2938391](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2938391)
125. Price WN. Medical Malpractice and Black-Box Medicine [Internet]. 2017 [cited 2019 Aug 9]. Available from: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2910417](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2910417)
126. Heather McEwen MM. NEOMED Library: Evidence-based Medicine I Course Guide (Archived): Introduction. [cited 2019 Aug 9]; Available from: [https://libraryguides.neomed.edu/EBM\\_I\\_Course](https://libraryguides.neomed.edu/EBM_I_Course)
127. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digit Med* [Internet]. 2019 Dec 11 [cited 2019 Oct 14];2(1):14. Available from: <http://www.nature.com/articles/s41746-019-0090-4>
128. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital Therapeutics: An Integral Component of Digital Innovation in Drug Development. *Clin Pharmacol Ther* [Internet]. 2018 Jul [cited 2019 Oct 14];104(1):72–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29377057>
129. Alami H, Gagnon M-P, Fortin J-P. Digital health and the challenge of health systems transformation. *mHealth* [Internet]. 2017 Aug 8 [cited 2019 Oct 14];3:31–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28894741>
130. 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* [Internet]. 2018 Dec 16 [cited 2019 Aug 9];1(1):26. Available from: <http://www.nature.com/articles/s41746-018-0034-4>
131. Healthcare Transformation Lab's Impact Report 2018 | Healthcare Transformation Lab at Massachusetts General Hospital [Internet]. [cited 2019 Aug 9]. Available from: <http://healthcaretransformation.org/2019/03/05/htl-impact-report-2018/>
132. Digital Health Study Physicians' motivations and requirements for adopting digital clinical tools. 2016.
133. The Johns Hopkins Hospital Launches Capacity Command Center to Enhance Hospital Operations - 10/26/2016 [Internet]. [cited 2018 Oct 14]. Available from: [https://www.hopkinsmedicine.org/news/media/releases/the\\_johns\\_hopkins\\_hospital\\_launches\\_capacity\\_command\\_center\\_to\\_enhance\\_hospital\\_operations](https://www.hopkinsmedicine.org/news/media/releases/the_johns_hopkins_hospital_launches_capacity_command_center_to_enhance_hospital_operations)
134. The Dark Secret at the Heart of AI - MIT Technology Review [Internet]. [cited 2017 Dec 23]. Available from: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>
135. EUR-Lex - 32016R0679 - EN - EUR-Lex [Internet]. [cited 2018 Nov 3]. Available from: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679>

136. The National Artificial Intelligence Research and Development Strategic Plan.
137. The use of algorithms in decision making: RSS evidence to the House of Commons Science and Technology Select Committee inquiry. [cited 2017 Dec 28]; Available from: <http://www.rss.org.uk/Images/PDF/influencing-change/2017/RSS%20evidence%20on%20the%20use%20of%20algorithms%20in%20decision%20making%20April%202017.pdf>
138. AMA Passes First Policy Recommendations on Augmented Intelligence | American Medical Association [Internet]. [cited 2018 Oct 15]. Available from: <https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence>



# APPENDIX

&

LIST OF PUBLICATIONS

ACKNOWLEDGEMENTS

AUTHOR AFFILIATIONS

ABOUT THE AUTHOR



## LIST OF PUBLICATIONS

**Paranjape K**, Schinkel M, Nanayakkara, PWB. Short Keynote Paper: *Mainstreaming Personalized Healthcare-Transforming Healthcare through new era of Artificial Intelligence*. IEEE J Biomed Heal Informatics [Internet]. 2020 Feb 10 [cited 2020 Mar 15];1–1. Available from: <https://pubmed.ncbi.nlm.nih.gov/32054591/>

Schinkel M, **Paranjape K**, Nannan Panday RS, Skyttberg N, Nanayakkara PWB. *Clinical applications of artificial intelligence in sepsis: A narrative review*. Comput Biol Med [Internet]. 2019 Oct 7;103488. Available from: <https://www.sciencedirect.com/science/article/pii/S0010482519303567>

Schinkel M, **Paranjape K**, Kundert J, Nannan Panday RS, Alam N, Nanayakkara PWB. *Towards understanding the effective use of antibiotics for sepsis*. Chest [Internet]. 2021 Apr [cited 2021 May 13];0(0). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0012369221007637>

Baron JM, **Paranjape K**, Love T, Sharma V, Heaney D, Prime M. *Development of a “meta-model” to address missing data, predict patient-specific cancer survival and provide a foundation for clinical decision support*. J Am Med Inform Assoc [Internet]. 2021 Mar 1 [cited 2021 May 13];28(3):605–15. Available from: <https://pubmed.ncbi.nlm.nih.gov/33260202/>

**Paranjape K**, Schinkel M, Hammer RD, Schouten B, Nannan Panday RS, Elbers PWG, Kramer MHH, Nanayakkara, PWB. *The Value of Artificial Intelligence in Laboratory Medicine*. Am J Clin Pathol [Internet]. 2020 Dec 14; Available from: <https://pubmed.ncbi.nlm.nih.gov/33313667/>

**Paranjape K**, Schinkel M, Nannan Panday RS, Car J, Nanayakkara, PWB. *Introducing Artificial Intelligence Training in Medical Education*. JMIR Med Educ. 2019;5(2):e16048. doi:10.2196/16048;103488. Available from: <https://pubmed.ncbi.nlm.nih.gov/31793895/>

**Paranjape K**, Schinkel M, Nanayakkara, PWB, Elbers PWG, Kramer MHH. Implementing Artificial Intelligence in Health Care: Data and Algorithm Challenges and Policy Considerations. Am J Biomed Sci Res. 2021 Jan 8;11(4):280–7. Available from: <https://biomedgrid.com/fulltext/volume11/implementing-artificial-intelligence-in-health-care-data-and-algorithm-challenges-and-policy-considerations.001645.php>

## AUTHOR AFFILIATIONS

Ketan Paranjape, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Michiel Schinkel, Section acute medicine, Department of Internal Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Prabath W.B. Nanayakkara, Section acute medicine, Department of Internal Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Rishi S. Nannan Panday, Section acute medicine, Department of Internal Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Paul W.G. Elbers, Intensivist | Deputy Chair ESICM Data Science Section, Department of Intensive Care Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Mark H.H. Kramer, Department of Internal Medicine, Section of Acute Medicine, VU University Medical Centre, Amsterdam, The Netherlands

Nadia Alam, Department of Internal Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

&

AUTHOR AFFILIATIONS

Bo Schouten, Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands.

Niclas Skyttberg, Department of Learning, Informatics, Management and Ethics, Health Informatics Centre, Karolinska Institutet, 171 77, Stockholm, Sweden

Josip Car, Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Richard D. Hammer, Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia.

Justin Kundert, Roche Diagnostics Corporation, North America, Indianapolis, Indiana, USA

Tara Love, Roche Diagnostics Corporation, Santa Clara, California, USA

Vishakha Sharma, Roche Diagnostics Corporation, Santa Clara, California, USA

Denise Heaney, Roche Diagnostics Corporation, North America, Indianapolis, Indiana, USA

Matthew Prime, Roche Diagnostics Corporation, Riehen, Basel Stadt, Switzerland

## **ACKNOWLEDGEMENTS**

Ever since I graduated from college in 1994, I have always had the dream of getting into academia to teach and do research. Following my graduate studies in computer sciences and electrical engineering, I enrolled for the PhD program at the University of Wisconsin – Madison in 1996. During one of the research symposiums, I had the opportunity to meet John Crawford from Intel Corporation who convinced me to join Intel Corporation to be a “chip architect” on a newline of microprocessors called Itanium. I decided to abandon my PhD but promised myself that I would eventually complete it, someday.

20 years later, on a business trip to Singapore I met Dr. Josip Car. I had then known Dr. Car for a few years having collaborated on a project when he was at the Imperial College. After our dinner conversation it became clear that it was time to get back on the PhD track.

Thank you, Dr. Car, for your friendship, mentorship, and partnership to help me succeed in this endeavor.

Through a fortuitous meeting with Dr. Louise van Galen who was doing her post doctorate with Dr. Car in Singapore, I learned about the work she was doing at the Amsterdam University Medical Center. We had an opportunity to collaborate on a paper on telehealth and the need to expand physicians’ communication competencies during telehealth.

&

This made me realize that developing and deploying technologies suitable for healthcare is one thing, but mainstreaming it for routine clinical use is another. The “end user” in this case the healthcare provider needs to be convinced on the clinical value technology brings to the diagnosis and treatment decision making. The technology has to blend in with their established workflows, guidelines and pathways. Finally, these users need to be trained on how to effectively use these new tools.

I decided to study how we can apply technologies such as AI to diseases like Sepsis and Cancer and mainstream them in routine clinical care. Along with generating clinical evidence to prove their efficacy, medical professionals will also have to be trained to use these technologies effectively. Having expressed my desire to work on this further and potentially do a PhD, I was introduced to Prof. Prabath Nanayakkara.

When I first googled him, to my utter surprise I found that he wasn’t just the head of acute medicine at VU medical center with a long list of world class publications, but was a prolific musician with songs on Spotify. Thank you Prof. Nanayakkara’s for your guidance, coaching and collaboration. He constantly challenged me to think from the vantage point of a clinician and how AI could blend into their routine workflows.

Through Prof. Nanayakkara, I was introduced to Dr. Michiel Schinkel and we decided to partner to work together on our PhDs. Michiel’s humility, work ethic, quest for knowledge

and healthcare expertise are the main reasons why I have succeeded in getting my PhD. I am also very grateful for the help from Dr. Hanneke Merten and Dr. Eva van den Ende who helped me with my university administrative work especially when I was 4000 miles away. Thank you Dr. Rishi Nannan Panday for your support and guidance through this journey.

Thank you to my promotor Prof. Mark Kramer and Co-promoter Dr. Paul Elbers for your support and guidance during my PhD program.

I would like to thank my collaborators who helped me during my research - Dr. Jason Baron and Dr. Denise Heaney on creating meta-model" to address missing data, predict patient-specific cancer survival and provide a foundation for clinical decision support; Justin Kundert for helping me with building machine learning models and Dr. Richard Hammer and Bo Schouten for identifying the value of AI in laboratory medicine.

It would be remiss of me if I did not share with you the vital role my family has played in my journey. Thank you to my parents and brother for shaping me into the person I am today. Your constant support and love are truly a blessing. I would like to thank my aunt Dr. Seema Joshi who has been a source of inspiration since my childhood and who encouraged me to pursue my PhD late in my career. I am here because of my grandmother Shakuntala Parkhi, Ajantha Parkhi and Dr. Sandeep Parkhi who helped me during a vital transition period in my life. I will forever be in your debt.

Finally, I am dedicating my PhD to the two most important people in my life, Anjali and Rishi Paranjape. Thank you for your love, support and encouragement through the many evenings and weekends I spent locked away in my room working on my papers. Without the two of you my life is incomplete and this is for you.

## ABOUT THE AUTHOR

Ketan Paranjape was born in Pune, India. Given his father's profession as a pilot in the Indian Navy, he grew up on multiple military bases enabling him to be exposed to multiple cultures, languages and ethnicities throughout his formative years. He attended 8 different schools and was consistently in the top 1% of his class along with pursuing competitive level Squash and Sailing. He got his undergraduate degree in Electrical Engineering from the University of Pune where he won the university gold medal for coming first with a GPA of 4.0.



Following a stint at Fujitsu developing software for clinical decision support, he attended the University of Wisconsin-Madison, USA to pursue a dual Master of Science degree in Electrical Engineering and Computer Sciences. After graduation he worked for 20 years at Intel Corporation in a variety of roles including head of engineering, head of product, chief of staff/technical advisor, head of commercial and finally as the general manager of the Health and Life Sciences business. During his tenure at Intel, he was a member of the US Health IT Standards Committee Precision Medicine Task Force, AAAS-FBI-UNICRI Project on Life Sciences and National and Transnational Security, International Telecommunication Union's Global Cybersecurity Working Group and World Health Organization's (WHO) Experts Working Group on eHealth.

He then spent 2 years as a Managing Director at Health2047, a silicon-valley startup funded by the American Medical Association developing and commercializing solutions for data liquidity, chronic care, productivity, security, and payments.

He is currently the Vice President of Diagnostics Information Solutions at Roche. The DIS business aims to harness the power of data, diagnostics and other critical information to support better clinical decisions. He is also on the roster of experts on Digital Health at the WHO.

A constant learner, he has completed the Stanford Executive Program from the Stanford Graduate School of Business, Healthcare Information Technology Management from the Harvard School of Public Health, Leading Digital Transformation in Healthcare from the Harvard Medical School and a MBA from the University of Oregon. He is currently pursuing a PhD at the Amsterdam University Medical Center, Vrije Universiteit in Amsterdam.

He resides in Indianapolis, Indiana, USA with his wife and son.

&

ABOUT THE AUTHOR

