University Of Bristol

Faculty of Engineering

Centre of Doctoral Training for Digital Health & Care

What antibiotic, when, in whom, and how much?

Precision antimicrobial prescribing in sepsis using individualized treatment effect inference

Edward Barker

Professor Andrew Dowsey
Dr Chris Bourdeaux
Dr Chris McWilliams
Dr Ranjeet Bhamber

October 22, 2021

Contents

1 Introduction **1**
- 1.1 Context . . . 1
  - 1.1.1 Health Data Research - United Kingdom (HDR-UK) . . . 1
  - 1.1.2 LABMARCS . . . 1

2 Clinical Need & Motivation **2**
- 2.1 Sepsis . . . 2
- 2.2 Antimicrobial resistance . . . 2
  - 2.2.1 Mechanisms of resistance . . . 3
- 2.3 Clinical practice . . . 3
- 2.4 Antimicrobial guidelines . . . 4
  - 2.4.1 Initiation . . . 4
  - 2.4.2 Evaluation . . . 5
  - 2.4.3 Narrowing therapy . . . 5
  - 2.4.4 Shorter therapy duration . . . 5
  - 2.4.5 Dosing . . . 5
  - 2.4.6 Multiple treatments at the same time . . . 5
  - 2.4.7 Conflict of interest . . . 6
- 2.5 Purpose . . . 6

3 Existing approaches **7**
- 3.1 The limits of traditional prediction . . . 7
- 3.2 Precision medicine . . . 8
  - 3.2.1 Observational datasets . . . 8
  - 3.2.2 Clinical trials . . . 8
- 3.3 Existing traditional approaches . . . 9

4 Research hypothesis **11**
- 4.1 Research Questions . . . 11
- 4.2 Outcomes of interest . . . 12
- 4.3 Causal inference for Individualised Treatment Effects . . . 12
- 4.4 Estimating response surfaces . . . 13
- 4.5 Including treatment effects in outcome models and handling bias . . . 14
- 4.6 Model selection . . . 15
- 4.7 Validation . . . 15
- 4.8 Domain adaptation . . . 15
- 4.9 Time-series data . . . 16
- 4.10 Trust (from decision makers and regulators) . . . 16
  - 4.10.1 Co-design . . . 17
- 4.11 Novelty . . . 17
  - 4.11.1 Technological novelty . . . 17
  - 4.11.2 Application novelty . . . 17

5 Methodology **17**
- 5.1 Main research problems . . . 17
- 5.2 Project plan . . . 17
- 5.3 Location / target population . . . 19
- 5.4 Inclusion / exclusion criteria . . . 19
  - 5.4.1 Ground truth of sepsis . . . 19

5.5 Data . . . 19
5.5.1 Systemwide Dataset . . . 19
5.5.2 MIMIC . . . 20
5.6 Data exploration . . . 20
5.6.1 Transition to Systemwide dataset . . . 24
5.7 Assess suitability of data . . . 24
5.7.1 Overlap assumption . . . 24
5.7.2 Unconfoundedness assumption . . . 25
5.8 Co-design . . . 26
5.8.1 Data-specific queries . . . 26
5.8.2 Prior knowledge of causal links . . . 27
5.9 Causal discovery . . . 27
5.10 Missing data . . . 27
5.11 Incremental approach . . . 28
5.12 Problem 1: Optimal antimicrobial in cross-sectional setting . . . 29
5.12.1 Small sample regime . . . 30
5.12.2 Large sample regime . . . 30
5.13 Problem 2: Optimal antimicrobial in temporal setting (timing) . . . 30
5.13.1 Time-dependent confounders . . . 30
5.13.2 Model choice . . . 31
5.14 Problem 3: Optimal dosing . . . 31
5.15 Evaluation . . . 31
5.15.1 Synthetic data . . . 32
5.15.2 Internal model validation . . . 32
5.15.3 Expert validation . . . 32
5.15.4 External model validation . . . 32
5.15.5 Deployment / Implementation . . . 33
5.15.6 Output / visualisation . . . 33
5.16 Extensions . . . 33
5.16.1 RCT data . . . 33
5.16.2 Pharmacometric and QSP approaches . . . 33
5.17 Responsible innovation . . . 33
5.17.1 Research governance . . . 33
5.17.2 Ethics . . . 33
5.17.3 Assessment and management of risk . . . 34
5.17.4 Open Science . . . 34
5.18 Public engagement . . . 34
5.19 National importance . . . 34

6 Bibliography **35**

7 LABMARCS **43**

1 Introduction

1.1 Context

1.1.1 Health Data Research - United Kingdom (HDR-UK)

This project is aligned with work conducted by members of the HDR-UK South West Better Care partnership team. The aims of the better care program are to build and scale data-driven learning healthcare systems across the UK to improve care and achieve better outcomes for patients, the wider public, and the health and care system.

The partnership is made up of several "Better care loops". There are two sub-projects relevant to my research, and they concern antimicrobial resistance and a Personalised National Early Warning Score.

The first project will develop of a Personalised National Early Warning Score (PNEWS). This will improve on the predictive performance of the National Early Warning Score with machine learning to personalise the score based on patient sub-types and by exploiting dynamic features extracted from the time series of electronic observations. The utilisation of the Systemwide dataset will allow one to learn better patient sub-types by using the patients' entire histories of interaction with the BNSSG health and care system.

The other project concerns the use of patient-specific antimicrobial history and the development of an antimicrobial resistance (AMR) risk model.

This PNEWS score will be utilised to identify a subgroup of patients with infection that are at risk of sepsis, as well as being incorporated as a covariate in the precision antimicrobial prescribing model. The patient-specific antimicrobial history and AMR risk model will be incorporated into my work to inform antimicrobial choice and predict outcomes.

1.1.2 LABMARCS

During the pandemic, no HDR-UK projects were allowed to move forward apart from COVID related projects. HDRUK refocused on COVID, and as part of this I did a large amount of work for a study called LABMARCS - Laboratory Markers of COVID-19 Severity - Bristol Cohort. The objectives were to:

1. Explore independent associations of routinely measured physiological and blood parameters at hospital admission with disease severity
2. Develop a prediction model for severe COVID-19 outcomes at 28 days in order to generate a simple severity scoring system so that the frontline clinicians can prioritise resources to the most critically ill patients to improve their outcome.

The laboratory markers in this study overlap considerably with those related to sepsis, and as will be demonstrated in the methodology section, this research is highly relevant to my thesis. The setting of the research, as well as the exploration and nature of the data are all valuable insights that will be transferable to the Systemwide dataset; the main dataset of interest in this project. Please see the appendix for the draft paper.

Approximately 6 months were spent working on COVID research, and this was followed by a suspension of studies for 3 months due to medical reasons. The timeline of both the LABMARCS research and the suspension are outline in the project plan.

2 Clinical Need & Motivation

2.1 Sepsis

Sepsis is an inflammatory immune response to an infection and can be a life-threatening medical emergency. It happens when an existing infection from places such as the lungs, skin and abdominal organs trigger a chain reaction that can cause damage to the body's tissues and organs. Severe sepsis is sepsis that can damage organ function or reduce blood flow [1]. Finally, septic shock is low blood pressure due to sepsis that does not improve after fluid replacement [1]. Both septic shock and severe sepsis are forms of sepsis, and we are interested in the more general definition of sepsis in the context of this project.

The Global Burden of Disease Study gave an estimated 48 million cases of sepsis worldwide in 2017, with 11 million sepsis-related deaths, a 20% mortality rate, which accounted for almost 20% of all global deaths [2]. However, it should be noted that the etiology and prevalence of sepsis in the UK is very different from that globally; in low-income countries diarrheal disease is the leading cause of sepsis, prevalence is higher and mortality is increased, whereas in the UK, the major cause of sepsis is lower respiratory infections [2].

In the UK, data from 133 hospital trusts shows that between 2014 and 2015 there were 55,171 hospital admissions for sepsis and 11,527 deaths, and between 2016-2017 there were 77,996 admissions and 15,851 deaths, an increase of 41% and 38% respectively. Although 70-80% of sepsis cases occur in the community, the vast majority of sepsis-related deaths occur in a hospital setting, with only 7% of sepsis related deaths not occurring in hospitals [3]. This increase has been attributed to an ageing population that is more vulnerable to infections that lead to sepsis, and antibiotic resistance [3].

Sepsis is a complex disease with a heterogeneous presentation, depending primarily on the identity of the pathogen causing it, the portal of entry, and the immunocompetence of the patient. The pathogen types causing sepsis can be bacterial, viral, or fungal, with the majority being bacterial.

The standard treatment for sepsis is antimicrobials. These are agents which are capable of destroying or inhibiting the growth of microorganisms, be they bacterial, fungal, viral. Antibiotics are those antimicrobials that are only active against bacteria. In the context of this report and project, we will talk about antimicrobials, and antimicrobial prescribing.

It is important to note at this point that there are a multitude of factors to consider with antimicrobial treatment, each of which is pertinent to two issues:

- The amount of antimicrobial used & the length of course
- The outcomes of patients.

Here I will go on to explain existing antimicrobial resistance, antimicrobial guidelines with respect to antimicrobial resistance, actual clinical practice, the points during the patient journey at which these decisions are made, and how these decisions affect antimicrobial use and patient outcomes. Finally, I will discuss existing clinical decision support systems to address precision prescribing

2.2 Antimicrobial resistance

Antimicrobial resistance occurs when microbes evolve mechanisms that protect them from the effects of antimicrobials. Antibiotic resistance is a subset of antimicrobial resistance, and it refers specifically to bacteria that become resistant to antibiotics. These more resistant

microbes are harder to treat, may require higher doses of antibiotics, and may require different types of antibiotic that are more toxic [4].

Antimicrobial resistance is a significant problem. Rising levels of resistance and a lack of new drugs in development means that current antibiotics must be used cautiously. Antimicrobial resistance causes 700,000 deaths worldwide, and this is predicted to rise to 10 million annually by 2050, with an accompanying cost of up to $100 trillion [5]. An annual Public Health England report shows that there were an estimated 61,000 an-tibiotic resistant infections in England during 2018, a 9% rise from 2017 [6]. Antibiotic-resistant bloodstream infections, a hallmark of sepsis, rose by a third(34%) between 2014 and 2018.

It is in our interest to take antimicrobial resistance into account when considering antimicrobial treatment plans, as there is something of a time limit on their efficacy; with the growing amount of resistance and with limited new antimicrobials showing promise in the near future. The clinical pipeline of new antimicrobials is "dry". "In 2019 WHO identified 32 antibiotics in clinical development that address the WHO list of priority pathogens, of which only six were classified as innovative" [4]

In the absence of promising new antimicrobials, it becomes more important to use existing antimicrobials conservatively, and only when necessary. It is our intention to engineer a clinical decision support system that gives clinicians the tools, knowledge and confidence to give antimicrobials only when they are needed, in the correct quantities, at the right time.

2.2.1 Mechanisms of resistance

2.3 Clinical practice

Current clinical practice often calls for treatment of sepsis with broad-spectrum antibiotics even before taking blood cultures to confirm the presence of bacteria1. Identification of the pathogen type happens when a sample of blood or other bodily fluid such as cerebrospinal fluid or urine is taken from the patient with suspected infection. However, it can take many days for definitive pathogen identification and for antimicrobial susceptibility testing [7]. Initial sepsis treatment bundles often occur before both pathogen identification and antimicrobial susceptibility testing, with a presumed bacterial infection and the administration of antibiotics. Here we find that the suitability of antibiotic treatment is not addressed.

If one gives broad-spectrum antibiotics without proper consideration of patient factors, patient prior exposure, and considering the source of infection in the body, then this may not be sufficient if the correct pathogens are not covered. Appropriate antimicrobial choice has shown to be one of the better predictors of survival [8, 9, 10, 11, 12], can reduce length of stay in the hospital, and reduce associated healthcare costs [7], so it is crucial that a more systematic approach is undertaken that takes all of these factors into account.

There is a clear need not only to identify the pathogen type of the infection, but it's associated antimicrobial resistance. However, local phenotypic resistance data is currently underutilised. This data is also lacking to clinicians. Data from primary care, secondary care, pharmacies, and laboratories is often kept separately, and this information is often unavailable at the time of treatment. Recent efforts have been made to integrate care records with examples such as Connecting Care in the Bristol, North Somerest and South Gloucestershire (BNSSG) Clinical Commissioning Group (CCG) [6], where records from multiple sources are able to be viewed from a single portal. However, whilst this information can be seen from the same portal, the disparate datasets are not currently in a format whereby they can undergo secondary data analysis.

2.4 Antimicrobial guidelines

The emergence of antimicrobial resistance has therefore lead to the creation of antimicrobial stewardship guidelines; this is an organisational or healthcare-system-wide approach to promoting and monitoring judicious use of antimicrobials to preserve their future effectiveness.

This report and project will make use of the recommendations given in the analytic review "Antibiotic Use in the Intensive Care Unit: Optimization and De-Escalation", which is a summary of the current literature on antimicrobial use in the ICU applying antimicrobial stewardship strategies [13].

Below is a diagram summarizing the treatment decision points in a patient's journey. For each decision point, I will explain the current clinical practice, the relevant antimicrobial prescription recommendations from the aforementioned review, and whether or not they are in-line with recommendations. Finally, the key research questions to be addressed will be listed, informed by the guidelines, and notes given on how this influences project development.

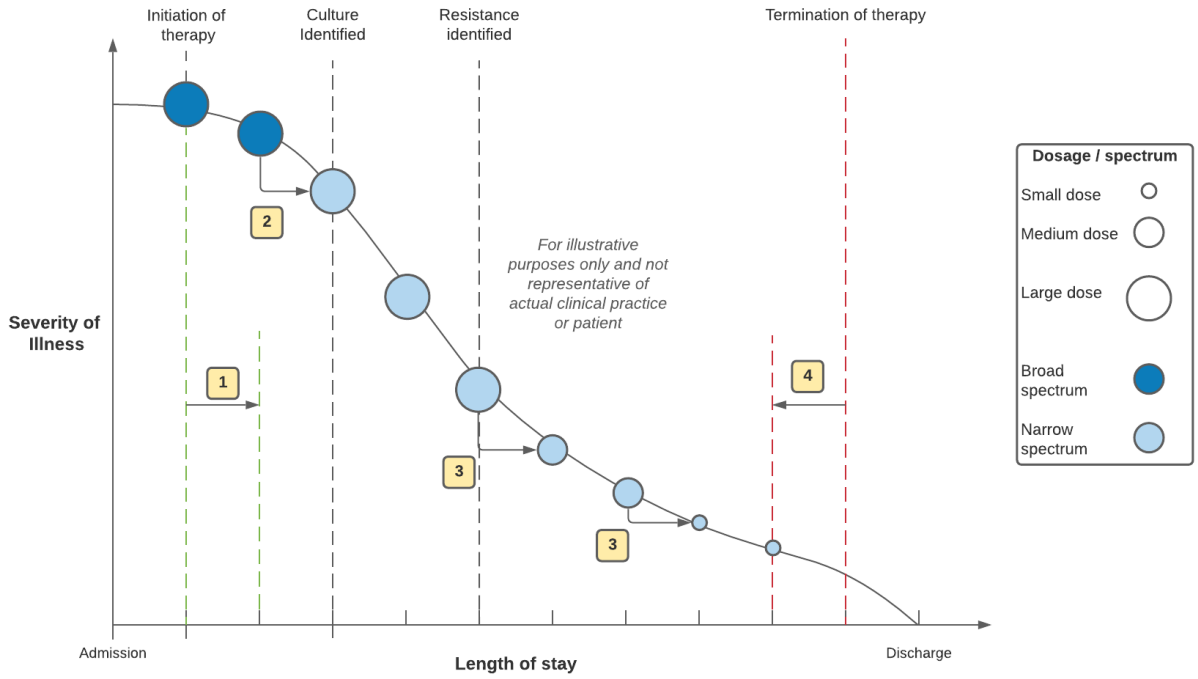


Figure 1: (1) Initiation of treatment, (2) Evaluation and narrowing therapy (3) Dosing, (4) Shorter therapy duration

2.4.1 Initiation

The identification of sepsis as early as possible is crucial in treating the disease, with earlier recognition of sepsis leading to better outcomes including decreased mortality. The earlier antibiotics are administered, the lower the mortality of the sepsis patient, and the less time they spend in the intensive care unit. The International Guidelines for Management of Sepsis recommend the administration of antibiotics within 1 hour of recognition of sepsis, a goal which the majority of hospitals are not currently achieving. Liu et. al. studied a population in the USA and found that the adjusted odds ratio for hospital mortality based on each hour of delay in antibiotics was 1.09 for each elapsed hour between identification and antibiotic administration [14].

2.4.2 Evaluation

There is a need to address a patient's ongoing reaction to an antibiotic treatment in real-time such that de-escalation and changes in treatment can be made as and when new information becomes available. This is an important consideration for our clinical decision support system; it must necessarily be a continuous prediction model able to operate on varying numbers of covariates. Treatment decisions are often reviewed on ward rounds on a daily basis, and one can consider a continuous prediction model that works in discrete time steps, though continuous should not be ruled out.

2.4.3 Narrowing therapy

Referring to Label 2 in Figure 2, a clinician may have the option to narrow the spectrum of an antimicrobial during the course of treatment. Many studies have shown narrowing antibiotic therapy does not adversely affect patient outcomes [15, 16, 17]. It should be noted that there is only limited evidence showing deescalation stops the development of resistant organisms, but I believe it is important to take this into consideration.

2.4.4 Shorter therapy duration

Therapy duration is important to consider here. Here we have to options for shorter duration:

- A later initiation of treatment
- An earlier termination of treatment

Label 1 of Figure 2 refers to a later initiation of treatment, and Label 4 refers to an earlier termination of treatment. As mentioned earlier, initiation of therapy as soon as possible after a sepsis infection is determined is crucial for increasing survival outcomes and so delaying treatment may well not be advisable. Nonetheless, many recent clinical studies and meta-analyses on common infectious diseases states have shown that shorter therapy is as efficacious as longer therapy [18, 19, 20], and so consideration of an earlier termination is really important with respect to our clinical decision support system.

2.4.5 Dosing

Label 3 of Figure 2 refers to the consideration of a dosage decision. It should be pointed out that this diagram is not representative of how actual care works, and dosage is not necessarily changed during treatment. Dosage is particularly challenging in the ICU, and response can be highly individualised. The amount of antimicrobial available at the site of infection can be influenced by the absorption, distribution, metabolism and elimination of the antimicrobial [**pea2005cp**]. For example, larger doses may be required for patients with pleural effusions (accumulation of fluid around the lungs)

Whilst choice of dosage bears less relevance to antimicrobial stewardship, it's certainly an important consideration for survival outcomes. A key conclusion to draw from this is the highly individualised nature of patient response to treatment in the context of antimicrobial prescribing. Our clinical decision support system must be highly individualised.

2.4.6 Multiple treatments at the same time

With the growing concern for gram-negative resistance and reduced susceptibilities, utilizing 2 antibiotics with different mechanisms of action for empiric treatment of infection will likely

become more important, especially in areas of higher bacterial resistance [21, 22]. This is crucial, as it means any clinical decision support system must be able to consider not just the best antibiotic of many, but the best combination of one or more treatments.

2.4.7 Conflict of interest

These guidelines exist because current clinical practice does not sufficiently address the steps needed to be taken to reduce the development of antimicrobial resistance. Clinicians are under many pressures and there are complex decision making environments that cause deviation from the guidelines.

Across the NHS, there are many factors at play, and competing priorities. There is a conflict of interest here for the clinician between the attempt to maximise the likelihood of survival for the individual patient, and benefit to all potential future patients through responsible antimicrobial use and decreased antimicrobial resistance. Such an abstract cost-benefit analysis is both not feasible and also difficult to visualise. It is important that our clinical decision support system is able to quantify the cost of antimicrobial use in terms of resistance, cost, side effects, and survival outcomes.

The irony here however is that adherence to many of the antimicrobial guidelines can and does increase survival outcomes for patients. These potential benefits need to be calculated and compared, and robust reasoning needs to be given behind such decisions.

Sepsis is not necessarily the cause of death for many of us, it is often preceded by other conditions. It is however the end of the pathway for a significant portion of people. Furthermore, the majority of sepsis patients end up in the hospital and the ICU. In this context, we both have access to a large amount of clinical information from electronic health records, and a problem that is opportune to the benefits of machine learning.

Below we give the purpose of the clinical decision support tool, the research questions to be answered, and the outcomes of interest. These research questions are informed by the context of the treatment; the ICU, the disease itself; sepsis, and current treatment guidelines and their adherence (or lack thereof).

2.5 Purpose

Information generation in health care is growing very quickly and outstripping the capacity of human cognition to adequately manage. Allowing clinical decision support systems to aid in decision making is a way to manage the limits of human cognition and make sense of the growing complexity of information. This work will develop a clinical decision support system that will make antibiotic prescribing recommendations for septic patients, using machine learning methods with their electronic health record information. It will be:

- Able to reduce antimicrobial use through optimal antimicrobial selection, initiation and deescalation
- Able to improve patient outcomes such as reducing mortality and length of stay through precision antimicrobial prescribing
- Trustworthy for clinical adoption through
  - Robust validation and evaluation
  - Being explainable and interpretable

3 Existing approaches

3.1 The limits of traditional prediction

Most traditional data science activities can be divided into two scientific tasks, each with different methods and philosophies: description and prediction.

Description (and visualisation) is focused on summarising, describing and visualizing features of interest, and involves simple calculations and unsupervised learning. This enables us to asks simple questions such as: what happened? Who was affected? What was the occurrence of Y in people with X? In our context this might be looking at the occurrence and spread of antimicrobial resistant pathogens.

Prediction is concerned with pattern recognition and forecasting, and involves statistical modeling and supervised learning. This allows us to ask questions such as: what will happen? Who will be affected? Are people with X more likely to have Y? In our context this might be predicting severity of infection.

The difficulty with our problem of interest is that we are asking questions such as "What would happen if..?", "If we changed X, how would it change Y?". In our context it examples would be "What would happen if we increased the dosage of antimicrobial X?" or "If we changed antimicrobial X, how would it change outcome Y?". To expound upon this, if we are to look at observational data of ICU stays of sepsis patients, we will find that a particular patient z was given treatment x and had an outcome y associated with this treatment. The crucial point here is that we only have the factual information of the outcome y paired with treatment x. If we change the treatment to another, we do not have a corresponding outcome, which would be known as a counterfactual (counter to the fact). One cannot answer such questions without some form of causal representation, whereby the counterfactual would be inferred from the causal relationships present in the model. These questions fall under the field of causal inference and counterfactual prediction, the "third pillar" of data science, specifically Individualised Treatment Effect Inference (ITE Inference). Unlike description and prediction, causal inference (and by extension ITE Inference) requires external contextual knowledge. This is not a unique problem to causal inference, as often prediction models infer causality regardless, and so are beset by this fundamental problem of causal inference and more!

The key thing to be asked in response to this conundrum is: can one answer such "what if?" questions without using causal inference and it's complexity? Does the question of what antibiotic is best to prescribe necessarily entail a causal approach? My argument is yes. In order to address this, I will discuss:

- What is needed from a clinical decision support tool to address precision antimicrobial prescribing
- Current approaches to solving this problem
- Explain why traditional prediction modeling doesn't give us what we want with respect to the problem
- Show how this new approach would solve these problems, as well as the fundamentals of how the approach works.
- Finally, discuss the added benefits of this approach with regards to generalizability, explainability and interpretability, trust, and acceptance.

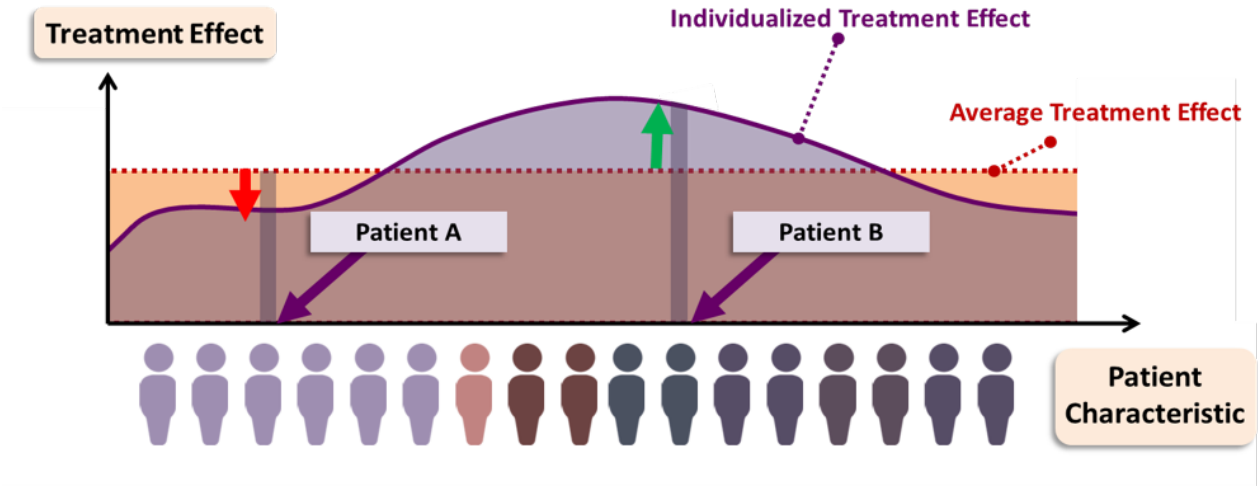


Figure 2: Average treatment effect vs individual treatment effect

3.2 Precision medicine

A major challenge in healthcare is determining whether or not a treatment influences or determines an outcome. For example, is there a survival benefit to prescribing a certain medication, such as the ability of a statin to lower the risk of cardiovascular disease?

Current treatment guidelines have been developed with the average patient in mind (on the basis of randomised control trials), but there is plenty of evidence that different treatments result in different effects and outcomes from one individual to another. For a given treatment, it is quite likely that only a small proportion of people will actually respond in a manner that resembles the average patient. INSERT EXAMPLE FROM ITE INFERENCE TUTORIAL.

There are two ways to determine whether a treatment works: observational datasets, and post-hoc analysis of clinical trials, and each method has its own strengths and weaknesses.

3.2.1 Observational datasets

Outside of guidelines; which apply to populations as a whole; or at best sub-populations, doctors can only learn from experience and time which treatments work best for each individual patient. There is no means by which this expertise can be shared on a population level in a way that allows insights on these treatment effects. As such, knowledge of these idiosyncrasies is shared in an unmethodical way, and will differ between clinicians.

There is an increasing availability of observational data, and this has encouraged the development of machine learning algorithms tailored for deducing treatment effects. It is crucial to note that observational datasets are prone to treatment assignment bias and other biases, and this will be explained in more detail later.

3.2.2 Clinical trials

Randomized Control Trials (RCTs) are the gold standard in medicine for comparing the effect of a new treatment to the current one. Through randomization, RCTs evenly distribute the known and unknown factors about patients into interventional and control groups, thereby reducing the potential bias due to confounding variables. However, RCTs may not always be the most practical option for evaluating certain treatments, as they are expensive and take a long time to implement.

Additionally, they are not always representative. This makes external validity a problem for RCTs as the findings can fail to generalise beyond the study population [23, 24, 25]. The narrow inclusion criteria in RCTs compared to the real world, where population restrictions with respect to disease severity, comorbidities, elderly patients, and ethnic minorities can be under-represented [26].

There is increasing awareness of this issue, but it is unlikely to be solved by RCTs and associated integrated and model-based analyses alone [27]. There is scope to add an adaptive element to clinical trials through the use of machine learning.

Notes on strengths and weaknesses in org-roam Draw from Peter Tennant talk on intro to causal inference - lots to be said about clinical trials. Try and make the point that all of the "weaknesses" of causal inference are really problems that already exist with current approaches, just wearing a different hat.

The research hypothesis here is to shift from a focus on the average treatment effect to individualised treatment effects by optimising the use of observational datasets and clinical trial design.

3.3 Existing traditional approaches

Alaa et al. [28], who use multi-task gaussian process (GP) experts to model the clinical trajectory of a patient given their subtype. Here there is significant overlap in methodology with the Better Care Partnership AMR project, and we will work closely with their team to adapt this GP modelling to our data and clinical context

In recent years, there has been a lot of research in MALDI-TOF(Matrix Assisted Laser Desorption Ionisation-Time Of Flight) Mass spectrometry for microbial identification and diagnosis [29]. This process is rapid, sensitive and economical in terms of reagent costs, but is potentially prohibitive in it's initial cost13. Additionally, although it solves the identification of the pathogen, there is still the problem of antimicrobial susceptibility testing. Whilst MALDI-TOF has been used for some forms of antimicrobial susceptibility testing [30], there are a plethora of examples of pathogens for which it is not able to differentiate between [31]. Additionally, MALDI-TOF does not take into account patient comorbidities, demographics, and other clinically relevant variables that influence the effect of the antibiotic, such as side-effects.

There has been work on clinical decision support systems (CDSS) for antimicrobial prescribing in the past. A systematic review on decision support tools to transform antibiotic management found that CDSS could improve antibiotic prescribing [32].

However, there are several shortcomings with these CDSSs, and these will be outlined hereafter.

Decision support systems for antibiotic prescribing generally fall into two categories: microbiology result independent prescribing, and microbiology result guided prescribing. Whilst microbiology result guided prescribing is useful for many types of infection such as urinary tract infections where blood culture time is not crucial, sepsis is particularly time-critical, and as such it makes sense to focus on studies that do not rely upon blood culture results.

Within result-independent prescribing, we can break the role of the decision support system into several sub tasks:

- Infection risk assessment
- Assessment of possible antibiotic profiles

- Choice of therapies

Infection risk assessment can be determined by probability calculators, often linked to provider-entered, patient-specific information. Work done using a simple personal digital assistant device used rudimentary pre-existing guidelines to guide decisions, and this was effective [33, 34] Significant work has been done in this area, specifically in predicting sepsis onset, and the risk of deterioration. As mentioned earlier, there is considerable heterogeneity in the population of those with sepsis, the symptoms of presentation appearing very differently depending on comorbidities, source of infection, pathogen type and demographics.

Assessment of possible antibiotic profiles can be provided through an interactive interface providing local cumulative antibiotic resistance data [35, 36].

Rudimentary approaches to a choice of therapies through electronic protocols and guidelines has been conducted, but these have no link to electronic healthcare records or laboratory information systems [35, 37].

It is the integration of these features that is key to an effective CDSS system, where the whole is more than the sum of it's parts.

TREAT uses a causal probabilistic network to evaluate patient details at presentation (demography, symptoms, background conditions, vital signs, test results) and outputs the probability of infection and its severity, source of infection, pathogen distribution, mortality and antibiotic coverage. A population-level 'ecological cost' for resistance is included [38]. This CDSS prescribed antibiotic treatment more frequently than clinicians (70% versus 57%), it used less broad-spectrum antibiotics, and cost less. This model could be used at any decision point during the course of antibiotic therapy, such that treatment could be de-escalated or changed. What is key to note about this model was that it only addressed empirical treatment, that is to say, treatment based upon experience. This model was static and built with expert guidance from published evidence. There was little opportunity to discover novel predictor variables.

A Bayesian approach has been proposed to factor in resistance rates over time [39]. However this is also an empiric treatment, and does not factor in new predictor variables.

Recently a retrospective observational study used routine clinical information for the XGBoost machine learning algorithm in predicting resistance to several common antibiotics. However, this performed no better than medical staff in the selection of appropriate antibiotics. It's limitations were that it selected patients with known infection with the pathogens of interest. Performance would need to be assessed prospectively with well-defined endpoints among patients with unconfirmed infection receiving a larger number of empirical choices.

The group integrated primary care and laboratory records in Israel across a 10-year period, linking 700,000 community-acquired UTIs with 5 million AM purchases to effectively predict resistance probability, and show associations with age, gender, care home residence and culminative AM use [40].

The use of sequencing to look at genomics data to assess properties such as identifying a species of an isolate, as well as it's resistance to antibiotics and virulence has been demonstrated by Didelot et. al. [41, 42, 43].

Treatment Complexity	Time Complexity	
	Static setting (S)	Temporal setting (T)
Binary treatment (B)	**(SB) In whom**? Who should receive antibiotic treatment or not?	**(TB) When** should treatment begin? **When** should treatment stop?
	(SBD) with accompanying dosage (**how much**)	**(TBD)** with accompanying dosage (**how much**)
One of many treatments (O)	**(SO) What** antibiotic is best to prescribe?	**(TO) What** antibiotic is best to prescribe and at what time to start and stop?
	(SOD) with accompanying dosage (**how much**)	**(TOD)** with accompanying dosage (**how much**)
Many of many treatments (M)	**(SM) What** combination of antibiotics is best to prescribe?	**(TM) What** combination of antibiotics is best to prescribe and at what times to start and stop?
	(SMD) with accompanying dosage (**how much**)	**(TMD)** with accompanying dosage (**how much**)

Figure 3: Research questions

4 Research hypothesis

4.1 Research Questions

The idea behind the format of these research questions allows several things:

Firstly, the complexity of the problem increases both with time and the number of treatments involved, as well as whether or not dosage is taken into account. As such, the problem will be approached as a progressive process, starting from the simplest model addressing the question of "Who should receive antibiotic treatment or not?" (SB), and finally ending with "Who should receive what combination of antibiotics, when, and how much?" (TMD). This progressive approach stops an overly-ambitious project being cut short, but also allows one to build upon previously created models. It also allows for comparison with other simple models already in the literature when complexity is low, and then the more complex questions are highly novel and have not been done in practice. These can of course all be compared with each other, and models from elsewhere in the literature, as the outcomes of interest will remain the same, and are listed below this section.

Secondly, one can pair the questions in this matrix with both the adherence to certain antibiotic guidelines, and also to address common errors in antibiotic use (happening in spite of antibiotic guidelines). The idea is that these various combinations solve different antibiotic guidelines.

In addition to this, there is an optional extended model with incorporation of pharmacokinetic and pharmacodynamic principles. If the three "axes" or combinations of the matrix were temporal complexity, treatment complexity, and dosage, the fourth would be the incorporation of pharmacokinetic and pharmacodynamic principles to increase cidality (cidality being the efficacy of the antimicrobial at killing the pathogen). This is particularly important when it comes to dosing.

The incremental development is listed in more detail in the methodology section

4.2 Outcomes of interest

- Day 7 and day 30 mortality
- Probability of recovery
- Length of hospital stay
- Amount of AM used
- Associated side effects of treatment
- Health and care system resource usage such as estimating the cost-effectiveness of using clinical prediction models to reduce AMR and AM in practice.

This study will look at patient-specific factors such as comorbidities and hospitilisations. Real-time information on infections, resistance and drug usage in other patients in a range of geographical proximities would enable further increases in resistance predictability [44]. This will be achieved through the integration of primary, secondary electronic healthcare records from the BNSSG Systemwide dataset with the Severn Pathology laboratory database; the diagnostic test provider for BNSSG. Patient precision will be improved through individually-linked animicrobial usage history and resistance data.

There is also an opportunity to incorporate genomics data. Ongoing research by Avison et. al. has defined rules for phenotypic resistance from whole genome sequencing to provide a "potential for resistance" score [45]. These potential for resistance scores will be incorporated into the model as additional covariates, aiding prediction.

We will use a Bayesian statistical learning approach as opposed to a pure machine learning strategy. This is because it would not be anchored to demonstrated statistical co-occurrence of resistance mechanisms. This will facilitate translation and use by clinicians and increase the clinical utility of the data generated.

Additionally, resistance rates and antimicrobial usage over time will be explicitly modeled, both patient-specific and population-level. This allows us to make predictions on future resistance rate pressures given the projected AM use, directly facilitating antimicrobial stewardship.

4.3 Causal inference for Individualised Treatment Effects

Estimating individual-level causal effects is usually done within the potential outcomes framework, which was introduced by Neyman in 1923 and then expanded by Rubin et al into a broader causal model. This framework is based upon observational data consisting of three parts:

- Patient features
- Treatment assignment

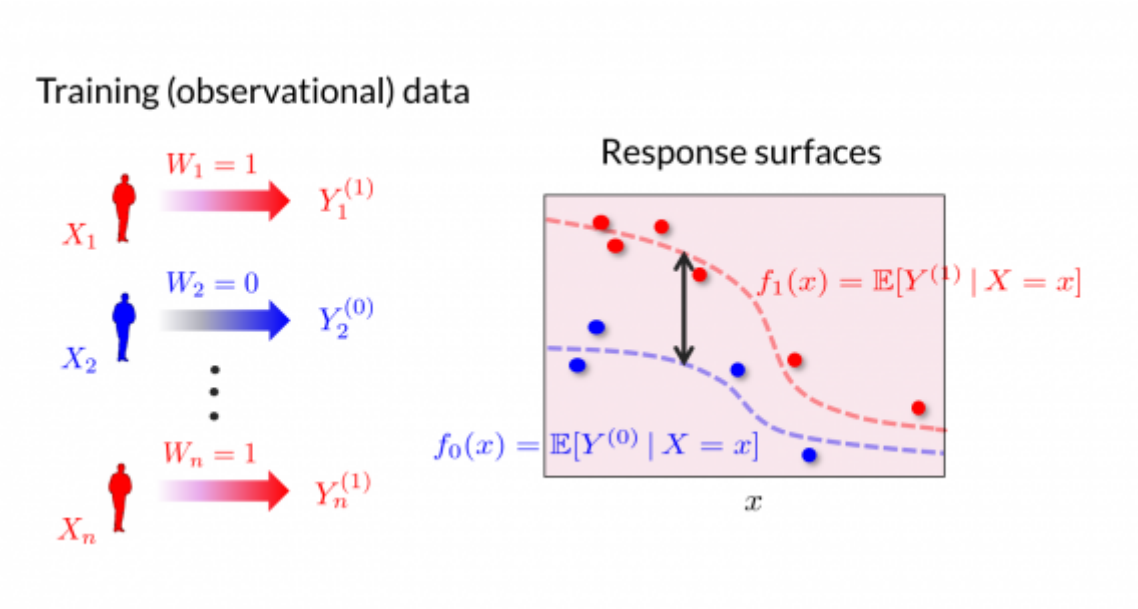


Figure 4: Response surface between two potential outcomes

- Outcomes

Although this framework is quite simple, the problem itself is quite complicated. There are a number of reasons for this, and we will address each in turn

1. We must work in the absence of counterfactual outcomes
2. Bias in observational datasets must be addressed
3. There is not a preferred way to include treatment indicators in outcome models

Much work has been done by the Van der Schaar lab to combat these problems, and this methodology draws heavily on their work, but builds upon their work with both methodological improvements, application to a new domain, and with the integration of more diverse data.

4.4 Estimating response surfaces

In the potential outcomes framework, every individual in the observational dataset has a number of potential outcomes: the subject's outcome under the application of various antimicrobials, and the subject's outcome when no antimicrobial is given. The treatment effect is the difference between the two potential outcomes, but since we only observe the "factual" outcome for a specific treatment assignment, and never observe the corresponding "counterfactual" outcome, we never observe any examples of the true treatment effect in an observational dataset. This is what makes the problem of individualized treatment effect inference fundamentally different from standard supervised learning (regression).

The majority of existing methods for estimating individualized treatment effects from observational data focus on the binary or categorical treatment settings and very few methods consider more complex treatment scenarios. However, it is often the case that antimicrobials have an associated dosage which requires us to estimate the causal effects of continuous-valued interventions.

We are potentially looking at the effects of high dimensional treatment with respect to antimicrobials, with their toxicity, side effects, and the potential for resistance. Work with more

complex treatment scenarios has been conducted with high dimensional organs for organ transplantation [46], and also on individualised dose-response estimation [47].

4.5 Including treatment effects in outcome models and handling bias

When modeling individualized treatment effects, we face further issues related to handling treatment bias in observational datasets, and a multitude of choices regarding approaches to handling treatment indicators when estimating patient outcomes.

Decision-making by doctors introduces bias into the data. When estimating individualised treatment effects, this assignment bias creates a discrepancy in the feature distributions for treated and control patient groups.

Modelling the treatment assignment, and its impact on the outcome, is a similarly complex proposition: several approaches exist, with the simplest being to split data into separate models (treated and untreated), or to use the assignment variable as a feature to augment the feature dimension.

A third solution, which has been adopted in a number of papers by the work of the Van der Schaar lab [48], is to learn shared representations, where the treatment assignment indexes these shared representations. This enables us to learn jointly across the treated and untreated populations.

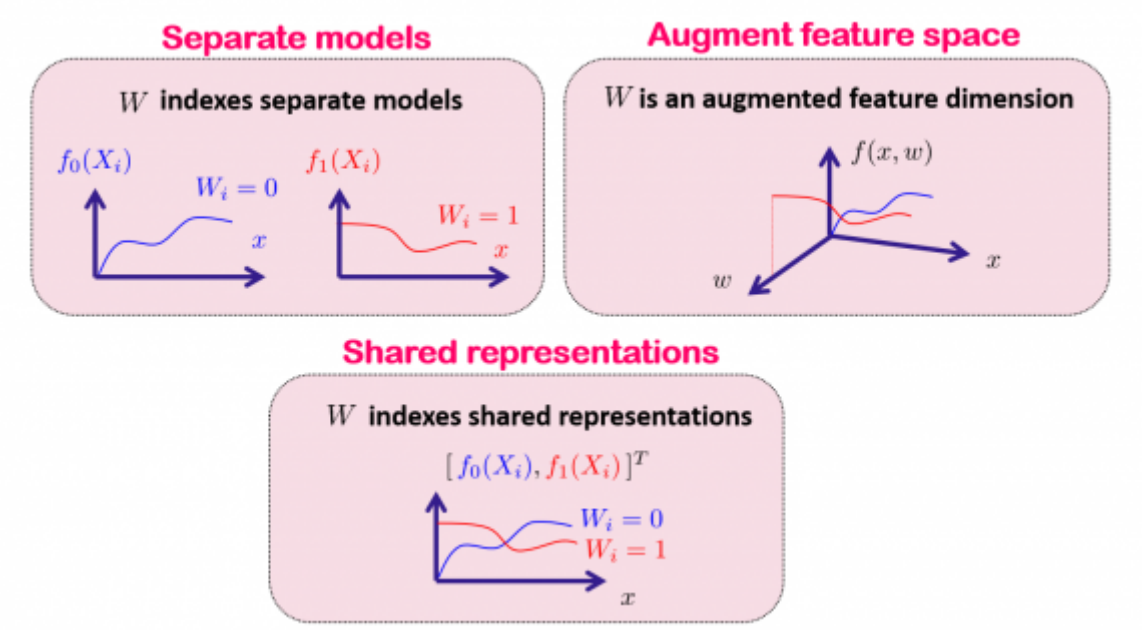


Figure 5

These methods all handle bias differently. The van der schaar lab developed a set of principles to guide the development of algorithms in this area, and broadly this depends on whether the development happens in a small sample regime or a large sample regime. This is discussed in detail in the methodology.

Finally, we are also able to look at the value of some continuous parameter associated with intervening, such as the dosage of antimicrobial to be given. Being able to better estimate individual responses to dosages would help us select treatments that result in improved patient

outcomes. Moreover, clinicians and patients will often need to consider several different outcomes (such as potential side effects); better estimates of such outcomes allow the patients to make a more informed decision that is suitable for them. Work by Bica et al looked at Estimating the Effects of Continuous-valued Interventions using Generative Adversarial Networks [47] that estimated dose-response curves for continuous interventions from observational data.

4.6 Model selection

Choosing "one best model" is impossible, since no single method will ever outperform all others across all datasets, so the challenge becomes selecting the best-performing model for each particular task and dataset. This is further complicated by the fact that we lack access to the counterfactuals and we cannot compute ground truth individualized treatment effects estimates to evaluate the model's predictions against. This is in contrast to predictive models, where one can use the mean squared error between the model's predictions and the ground truth label. The answer to this problem is to use automated machine learning (AutoML) to compare models and select the best model for the task at hand. In experiments applying an AutoML framework for individualized treatment effect inference (details of which are provided in the box below), the van der schaar found that the best model selected by the framework tended to significantly outperform other commonly-used methods.

4.7 Validation

In an ICML 2019 paper, entitled "Validating Causal Inference Models via Influence Functions," Alaa et al introduced a first-of-its-kind validation procedure for estimating the performance of causal inference methods using influence functions (IFs)—the functional derivatives of a loss function.

The procedure they introduced utilizes a Taylor-like expansion to approximate the loss function of a method on a given dataset in terms of the influence functions of its loss on a "synthesized", proximal dataset with known causal effects.

This automated and data-driven approach to model selection enables confident deployment of (black-box) machine learning-based methods, and safeguards against naïve modeling choices.

4.8 Domain adaptation

It is often the case that the observational data used to train a treatments effect model may come from a setting where the distribution of patient features is different from the one in the deployment environment, and this may be the case with our study, with the use of the MIMIC dataset from the USA, and the Systemwide dataset, which is for a Bristol population. Because of this, it is important to be able to also select models that are robust to these covariate shifts across disparate patient populations.

In a recent paper from the van der schaar lab, they proposed leveraging the invariance of causal structures across domains to introduce a novel model selection metric specifically designed for treatment effects models under the unsupervised domain adaptation setting. Experimentally, their method selects treatment effects models that are more robust to covariate shifts on several synthetic and real healthcare datasets [49].

4.9 Time-series data

While the majority of previous work focuses on the effects of interventions in a cross-sectional setting, observational data also capture information on complex time-dependent treatment scenarios, such as where the efficacy of antimicrobials changes over time, or where patients receive multiple antimicrobials administered at different points in time.

Estimating the effects of treatments over time therefore presents unique opportunities, such as understanding how sepsis evolves under different treatment plans, how individual patients respond to antimicrobials over time, and which timings may be optimal for assigning treatments, thus providing new tools to improve clinical decision support systems.

Estimating counterfactual patient outcomes over time is challenging due to the presence of time-dependent confounders in observational datasets. Time-dependent confounders are patient covariates that affect the treatment assignments and are themselves affected by past treatments.

To make this even more challenging, estimating the effect of a different sequence of treatments on the patient would require not only adjusting for the bias at the current step (in treatment A), but also for the bias introduced by the previous application of treatment B. Using standard supervised learning methods to estimate these treatment effects will be biased by the treatment assignment policy present in the observational dataset and will not be able to generalize well to changes in the treatment policy in order to generate counterfactuals. Approaches to handling time-dependent confounders are handled in the methodology.

The ability to accurately estimate treatment effects over time using machine learning allows clinicians to determine, in a manner tailored to each individual patient, both the antimicrobials to prescribe and the optimal time at which to administer them, given their observational history.

4.10 Trust (from decision makers and regulators)

It is important to consider the problem domain holistically during the development of a clinical decision support system. A CDSS can have an extremely high performance, but may be lacking in features that make it attractive for use in actual clinical practice. Eichler et al discuss requirements for making machine learning methods acceptable for decision makers and regulators [50].

- The importance of domain knowledge for providing the assurance that the unconfoundedness assumption actually holds in the observational dataset of interest.
- Expert knowledge is also needed to validate the model's estimate of the treatment effects.
- The patient population used to train the model must be representative of the patient population the model is to be deployed on.
- Clinicians need understanding of which patient features are the strongest drivers of treatment response, particularly when the patient data are high-dimensional.
- Uncertainty estimates are also beneficial for assessing the model's confidence in its predictions (if outcomes estimated by a model are highly variable, it then becomes difficult to trust the model safely and consistently for treatment decisions.

FAT forensics would be a good thing to mention here: a toolbox(https://fat-forensics.org/index.html) for interpretability, fairness, accountability and trust.

4.10.1 Co-design

Co-design in the context of designing a system that will be easy to use and trusted by clinicians, keeping all stakeholders in mind, and being realistic about implementation.

Moreover, collaborative research between medical practitioners and ML researchers is increasingly common [51]

4.11 Novelty

4.11.1 Technological novelty

- Estimating the individualised effects of time-dependent treatments with associated dosage
- Incorporation of pharmacokinetic and pharmacodynamic principles to optimise dosing
- Modeling combinations of treatments assigned over time using ITE inference

4.11.2 Application novelty

- Individualised treatment effect inference of antibiotics in the context of sepsis (novel domain)
- The use of the breadth of data from primary care, secondary care, local susceptibility data, and local pathology from the Systemwide dataset.

5 Methodology

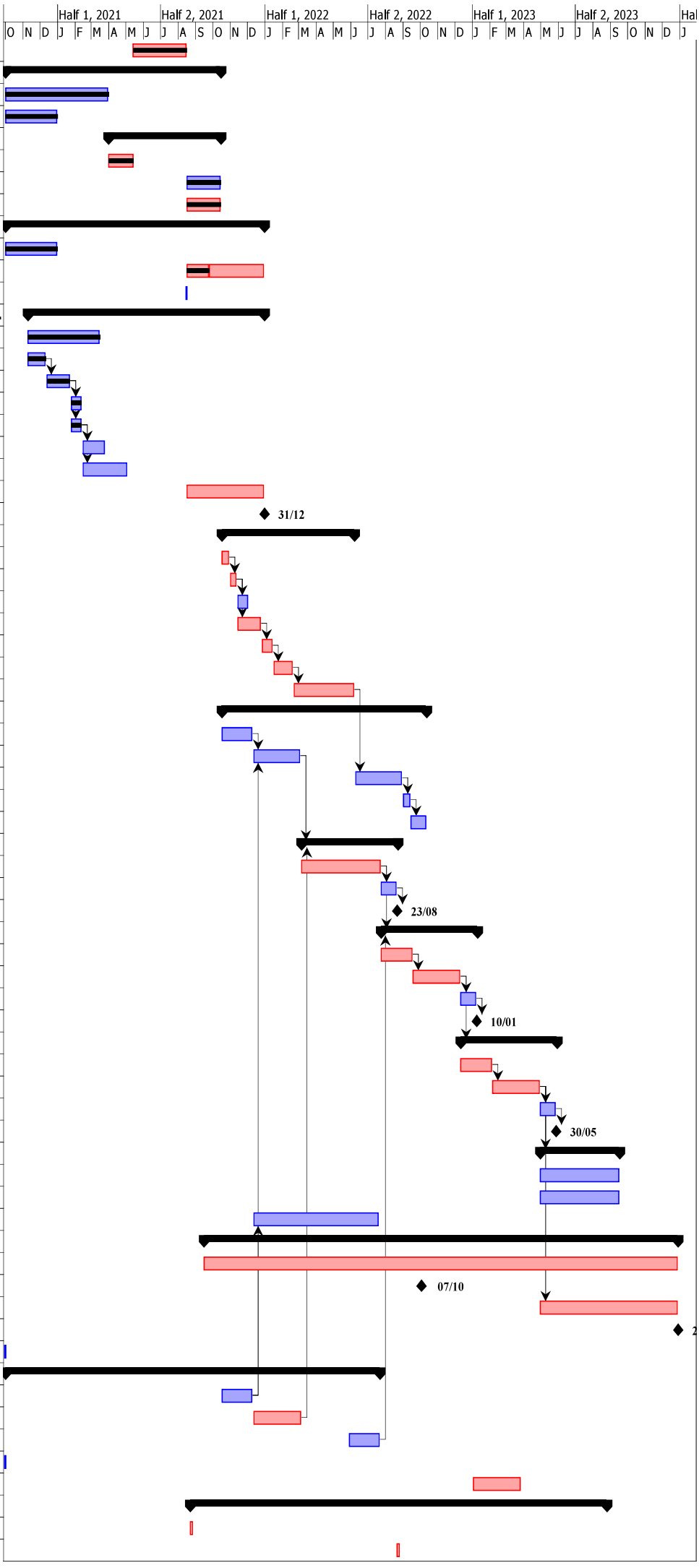
5.1 Main research problems

1. Optimal antimicrobial prescribing in the static setting
2. Optimal antimicrobial prescribing and timing in the temporal setting
3. Optimal antimicrobial dosing

These are the broad research questions that will be addressed within the context of the PhD. Here the methodology will be approximately chronological, with parallel steps shown in the project plan. We will discuss pre-processing and testing of key assumptions, introduce methodology specific to each of the three key problems, and finally discuss evaluation and possible extensions

5.2 Project plan

	Name	Duration
1	Suspension of studies	67 days
2	**Literature**	**272 days?**
3	Literature search	131 days?
4	Research proposal	67 days?
5	**Lit review**	**142 days?**
6	Literature review	32 days
7	Literature review	44 days?
8	Research plan	45 days?
9	**Training**	**327 days?**
10	Andrew Ng Machine Learning Course	67 days?
11	Mathematics for Machine Learning	100 days?
12	Causal Inference self-taught	1 day?
13	**LABMARCS study**	**299.875 da...**
14	Literature search	92.875 days?
15	Data cleaning	24.875 days
16	Data exploration	30 days
17	Data visualisation	15 days
18	Transformation	15 days
19	Imputation	30 days
20	Modeling	57 days
21	Paper writing	100 days?
22	Submission to journal	0 days
23	**Systemwide data**	**168 days**
24	Get CCG sponsor	10 days
25	BI Work Request form	10 days
26	Approval	15 days
27	Generate detailed data spec	30 days
28	Complete DPA schedules/tasking forms	15 days
29	UoB arranges DSPT specifics & data safe haven	25 days
30	PA prepares data and shares via MESH	78 days
31	**Data**	**258 days**
32	MIMIC data exploration	40 days
33	Construct broad ML pipeline	60 days
34	Transition to Systemwide dataset	60 days
35	Assess suitability of data	10 days
36	Make ready for cross-comparison	20 days
37	**Problem 1: Optimal antimicrobial (static)**	**121 days?**
38	Optimal in static setting	100 days
39	Paper write-up	20 days
40	Paper submission	1 day?
41	**Problem 2: Optimal antimicrobial + timing**	**121 days?**
42	Timing of single	40 days
43	Timing of one of many	60 days?
44	Paper write-up	20 days
45	Paper submission	1 day?
46	**Problem 3: Optimal dosing**	**121 days?**
47	Dosage of single	40 days
48	Dosage of one of many	60 days
49	Paper write-up	20 days
50	Paper submission	1 day?
51	**Extensions**	**100 days**
52	Multiple treatments	100 days
53	Incorporation of RCTs, QSP and PK/PD	100 days
54	Causal discovery / bias	159 days?
55	**Thesis**	**598 days?**
56	Ongoing writing	598 days?
57	Thesis completiton plan	5 days
58	Dedicated writing	174.875 days?
59	Thesis submission	0 days
60	Ethics	1 day?
61	**Co-design**	**472 days?**
62	Data-specific queries	40 days
63	Identification of confounders	60 days
64	Identification of time-confounders	40 days
65	Validation of treatment predictions	1 day?
66	Studentship	60 days
67	**Public involvement**	**525 days**
68	We the curious residency '21	5 days
69	We the curious residency '22	5 days



5.3 Location / target population

Patients in the NHS Bristol, North Somerset and South Gloucestershire CCG boundary

5.4 Inclusion / exclusion criteria

Inclusion criteria:

1. All adult patients in hospitals across Bristol with a positive diagnosis of Sepsis during a stay in the ICU

5.4.1 Ground truth of sepsis

There are many scoring systems and definitions for sepsis. None are perfect and many seek to measure similar variables. In February 2016, the sepsis-3 definitions were released and sepsis was defined to be a "life-threatening organ dysfunction caused by dysregulated host response to infection" [8]. Organ dysfunction, in turn, is defined as a change of 2 or more points in the Sequential (or Sepsis-related) Organ Failure Assessment (SOFA) score.

The National Early Warning Score (NEWS) uses several bedside measures including blood pressure and pulse rate, and the system was initially designed for paper chart use. This makes it a discrete system, not able to take into account changes over time. These measures are now recorded electronically. This score is nonspecific, and as such there are sepsis patients with low NEWS scores admitted to the Intensive Care Unit with septic shock. • SOFA has a poor sensitivity, leading to delays in sepsis identification [52] • SIRS has high sensitivity and could lead to over diagnosis, resulting in inappropriate antibiotic use [53]

Several other models have been developed for use in the ICU, including APACHE III, the Simplified Acute Physiology Score, and Mortality Probability Model II [54, 55, 56].

Ultimately the aim is to integrate with the PNEWS project to use their scoring system as the ground truth for sepsis diagnosis in the ICU.

All of these scores, as well as diagnostic codes, will be used to determine sub-populations who are considered to have sepsis, and who therefore are suitable for inclusion in the project. Use of multiple scores will allow for comparison with a diverse set of models used in the past, who use different diagnostic criteria.

Exclusion criteria

1. Paediatric patients (age <18 years)
2. Staff / healthcare worker house-hold contacts

5.5 Data

- The BNSSG Systemwide dataset
- The MIMIC-III and MIMIC-IV critical care databases

5.5.1 Systemwide Dataset

Information about antibiotic use and antimicrobial resistance is held within difference databases, such as General Practices, laboratories, pharmacies and hospitals, across the health system. We will analyse data from the Bristol, North Somerset and South Gloucestershire (BNSSG) Systemwide dataset of linked primary care, secondary care and laboratory records to optimise

antibiotic choice for patients based on their individual history and clinical characteristics, as well as their population's risk of resistance. We will collect pathology data including bacteriology, virology, mycology, haematology and biochemistry. We will exclude all staff members and their household contacts, and based on their age we will exclude all paediatric patients. Longitudinal data will allow us to identify transfer to the intensive care unit and mortality including the timelines.

The process of accessing the Systemwide dataset involves requesting variables of interest rather than doing exploratory data analysis. The use of the MIMIC dataset will inform variable choice for this dataset.

5.5.2 MIMIC

MIMIC-IV was released in 2021 and contains many new sources of data that make it suitable for use for our problem of interest. Electronic Medicine Administration Records (EMAR) are now present, which allows for precise timing of medication administration. Prescription information now has time instead of just date, enabling more granular, high resolution record of when prescriptions started/stopped. Finally, there is more information about dosing, and more information about prescriptions themselves.

The benefit of the MIMIC dataset is threefold. Firstly, MIMIC is a very accessible and freely available Intensive Care Unit dataset, and this has enabled it's use for benchmarking the performance of comparable machine learning methods. This will allow us to objectively compare the performance of our machine learning models against previous approaches

Secondly, the acquisition of the Systemwide dataset is an involved process and relies upon external partners. The timeline of it's delivery is very uncertain and as such it is important to make the best use of time available by working on an alternative dataset.

Finally, the use of domain adaptation will be explored, and the size of the MIMIC dataset leveraged to this end.

5.6 Data exploration

Approximately 8 weeks will be spent exploring the MIMIC dataset; determining our population of interest using the different diagnostic criteria, assessing which variables are of importance, making the data compatible with our machine learning pipeline, and taking into account data-specific considerations with respect to methodology choices.

This was an important part of the LABMARCS research project, and by no means is data exploration limited to this time period, it is just this period in which it will be the primary focus. Very similar approaches and use of the analyses and tools developed in LABMARCS will be utilised during the data exploration in both the MIMIC dataset and Systemwide.

We first considered demographics, and stratified by age and gender to get an idea of what kinds of people came to hospital with COVID.

Length of stay was an important consideration. The goal of the research was to identify laboratory markers of severity for COVID-19. As such, it was important that we considered short enough time ranges such that they would be clinically useful. If one needs to give a battery of tests over 14 days, many patients may have deteriorated by this point, and so a balance must be struck between the coverage of more patients and more data with a longer time period, and the cost of needing a long time to decide care provision.

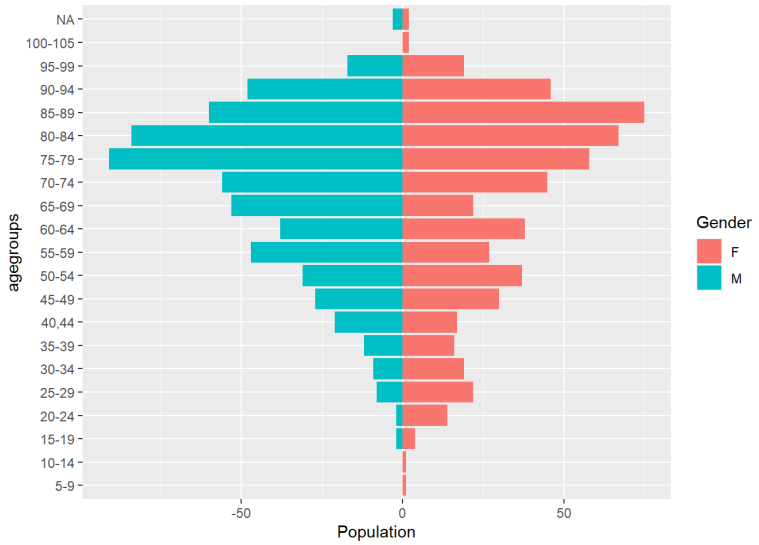


Figure 6: Stacked pyramid plot of LABMARCS population, stratified by age and gender

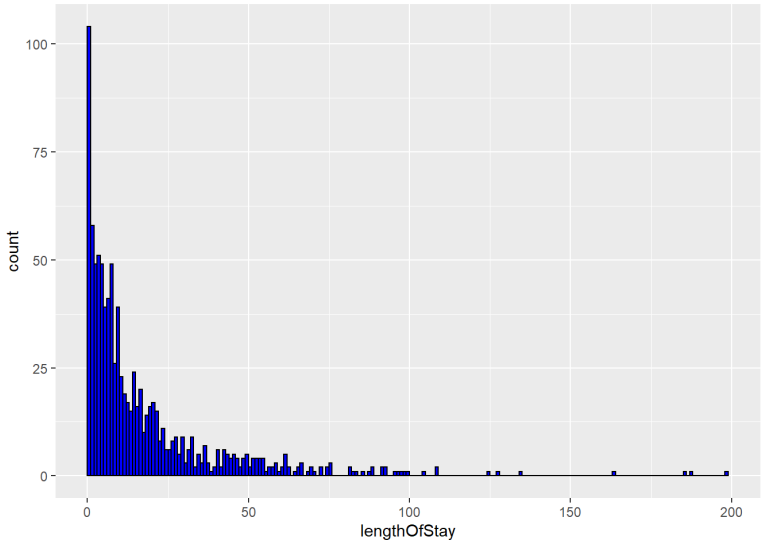


Figure 7: Frequency plot of length of stay for covid patients positive patients

It was also important to consider which patients went to the ICU, as this would indicate a more severe infection. We considered stratifying the population based on this, as both us and clinicians suspected that they may have very different clinical trajectories and therefore require different risk scores. The precision antimicrobial prescribing model may well differ in this regard as the stratification is handled by individual characteristics rather than manual separation. That being said, we may want to consider sub-populations depending on severity of sepsis, and this will be informed by co-design sessions with the clinicians.

We made variable centric plots for different sub-populations to see which variables differed from the reference range of "normal patients", as this would influence our choice of features to be included in the model. Our model ended up using reference ranges to determine what values were "bad" or "worse". This was complicated by the fact that it was not clear what was worse, below a lower bound or above an upper bound. This is another reason why conversations with clinicians is crucial.

Visualisation of patients will be used to gain an understanding of events during a patient's stay, missingness patterns, the pattern of sampling, as well as number of measurements per patient

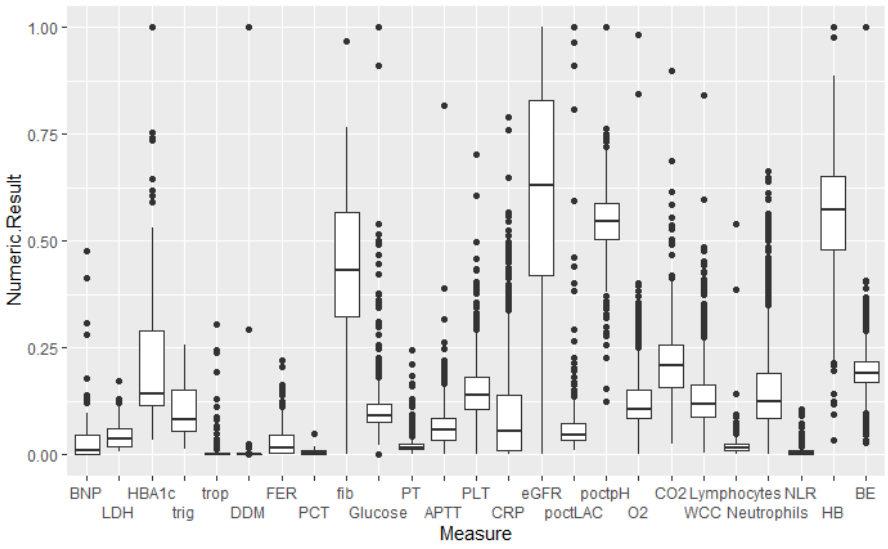


Figure 8: Laboratory measurement values on date of admission (normalised)

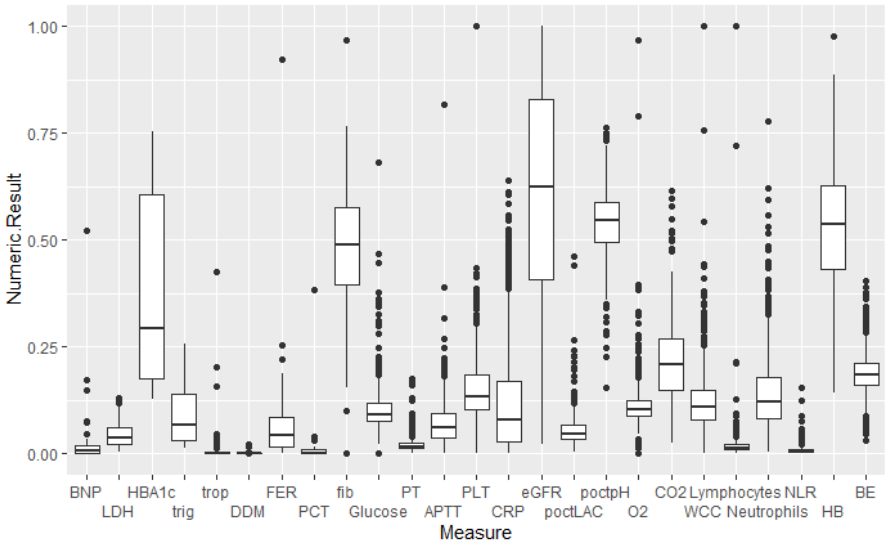


Figure 9: Laboratory measurement values on date of +ve covid test (normalised)

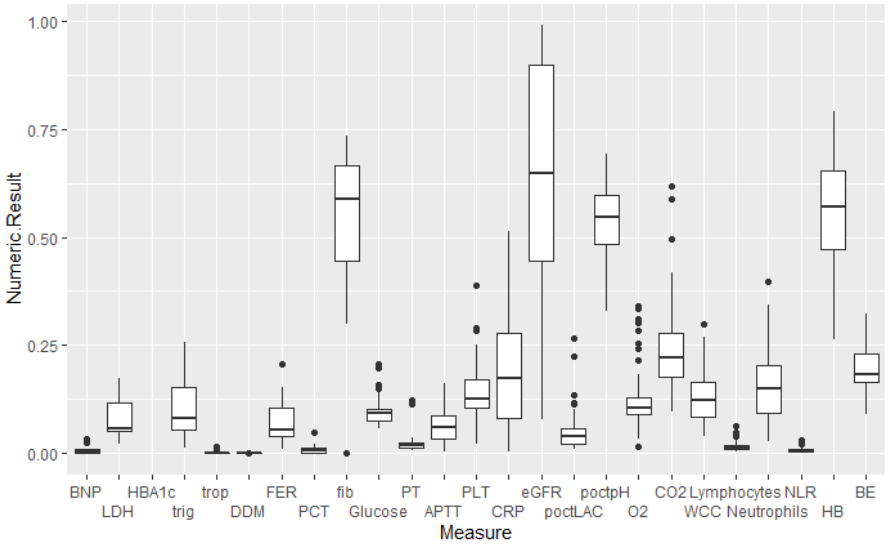


Figure 10: Laboratory measurement values on date of admission to ICU (normalised)

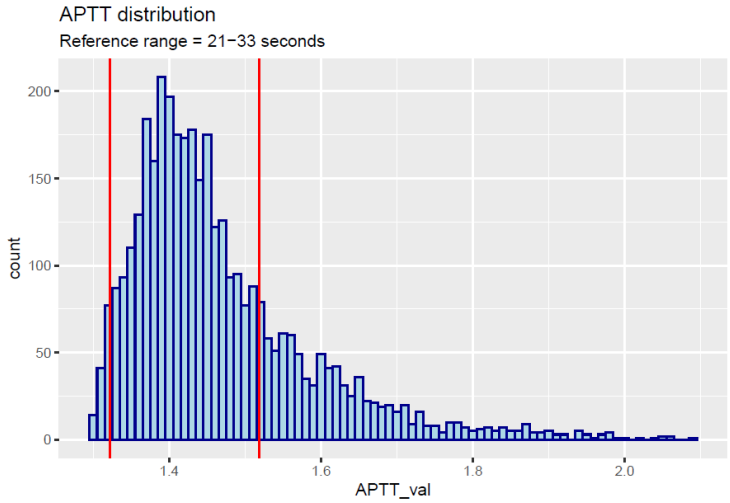


Figure 11: Histogram of Partial thromboplastin time value distribution

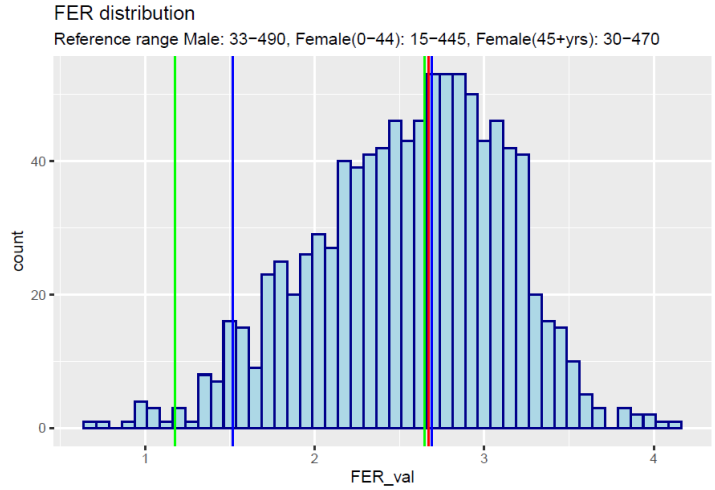


Figure 12: Histogram of ferritin distribution, demonstrating reference ranges specific to age and gender

and measurement frequency.

This will be achieved in a similar fashion to work on the LABMARCS dataset, where I visualised the patient stay to consider things such as the "alignment point" which we would consider as the start of our patient's stay.

Patient-centric visualisations were the easiest way to gauge any significant problems with the data. For instance, some patients were found to have discharge before admission, and this turned out to be a problem with the way the data was acquired. Some patients had multiple admissions when we only wanted to consider their stay relevant to the covid test date.

Although the disease of interest between LABMARCS and the main research project are different, there are data-specific problems that are common to both. Not only this, but the laboratory markers that we consider are very similar, as is the setting, with many COVID patients staying in the ICU. Work done on LABMARCS has been instrumental in finding common problems with health data, and also highlight the importance of continuous collaboration with health care professionals; helping spot problems that could not be found from the data alone.

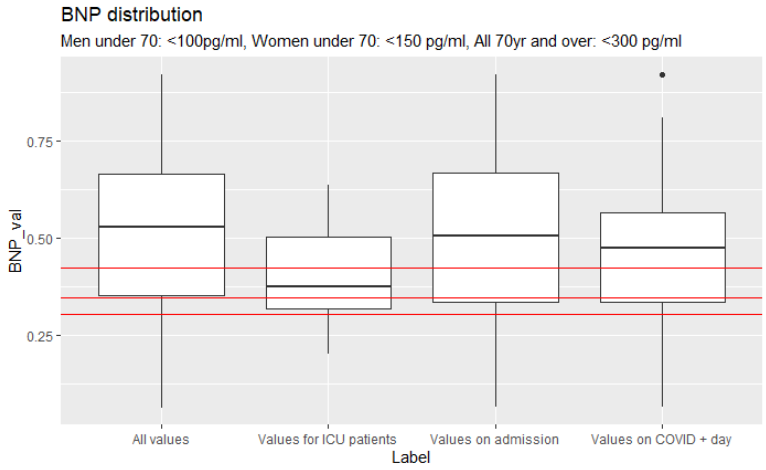


Figure 13: Box plot of B-type natriuretic peptide values, showing references ranges for different sub-populations

5.6.1 Transition to Systemwide dataset

We must ensure that the patient population used to train the model is representative of the patient population the model will be deployed on. Since development will start on MIMIC, there must be time set aside to compare the populations of these datasets, as well as looking at the commonalities in terms of variables. If for instance one wished to leverage the size of MIMIC for use in the Systemwide dataset, and the population of MIMIC was not representative of Systemwide, then additional work would need to be done to see if the causal inference models trained on MIMIC were robust to distributional shifts in the patient population. If the population in MIMIC was not representative of the Systemwide dataset and the model was not robust to changes, then early stages of the machine learning pipeline would have to be repeated. Approximately three months will be set aside for this transition.

5.7 Assess suitability of data

Causal inference using observational data make a few key assumptions common to all the problems of interest in this project; overlap, and no hidden confounders. If these assumptions fail to hold, then we can say that the observational data is not suitable for making personalised antimicrobial recommendations.

5.7.1 Overlap assumption

Overlap means that there is common support between the treatment groups. Put another way, this means that each patient has a non-zero probability of receiving each antimicrobial. If we do not observe treatment alternatives in similar contexts then it is not possible to make reliable treatment recommendations. As an example, if patients with diabetes always receive the same treatment then we cannot estimate the effects of other possible treatments on these patients. Overlap can be evaluated by working out probabilities of treatments for each patient. If these are bounded away from 0 and 1 for each patient, then there is overlap between the treatment groups. Overlap becomes more complicated when there are high dimensional features for the patients.

We will use the current state-of-the-art method for assessing overlap, developed by Johansson et al [57]. This identifies overlap regions by finding the minimum value sets of the observational distribution that are subject to coverage constraints by estimating treatment assignment

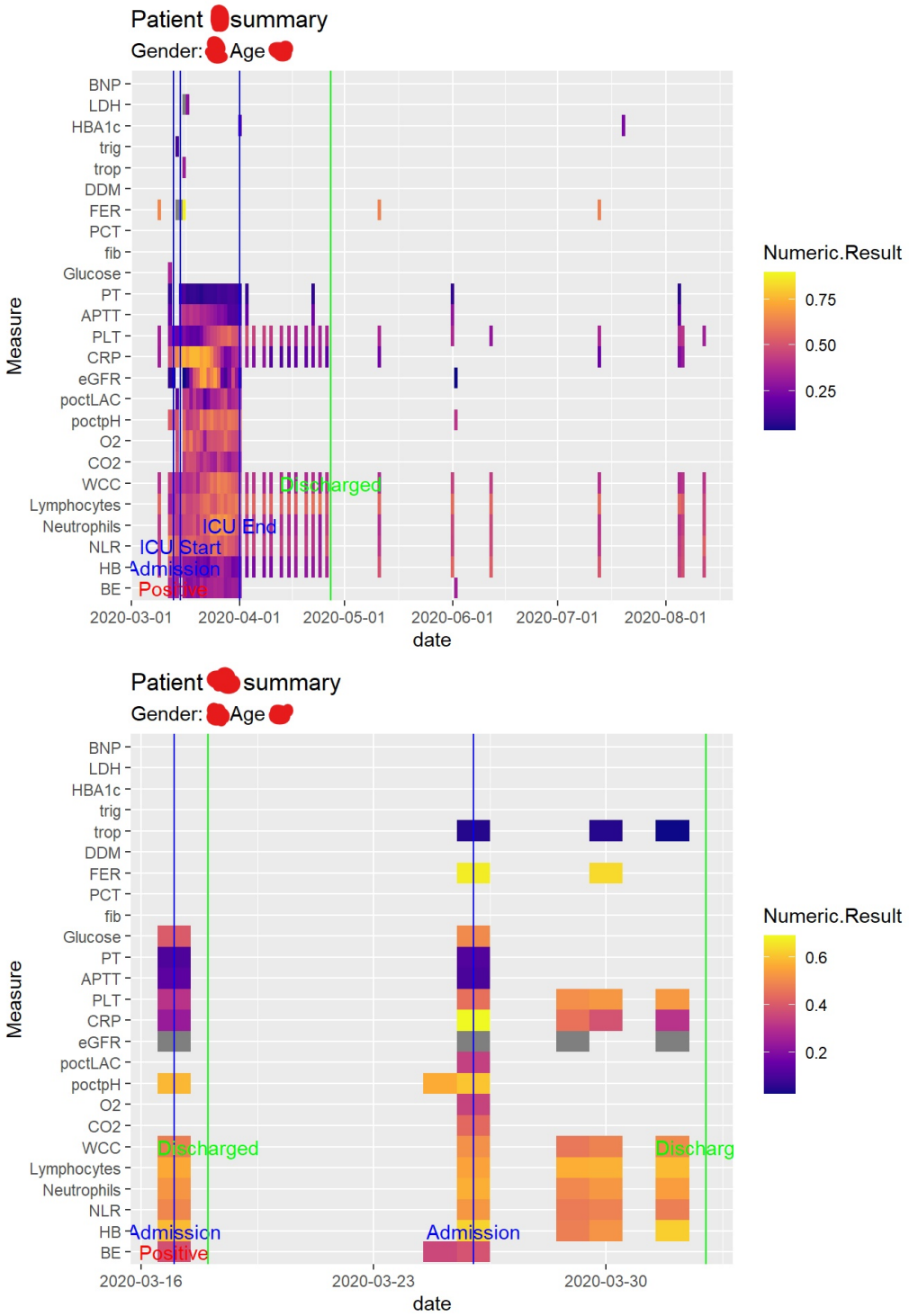


Figure 14: Visualisations showing when measurements were taken during a patient's stay, with their normalised value represented with a heat map. Dates of admission, discharge, +ve covid test and ICU start and end dates are also included.

propensities. The algorithm, called "OverRule", outputs a rule-based characterisation of overlap between treated populations.

5.7.2 Unconfoundedness assumption

Confounders in our context are all variables that affect both the treatment assignment and the outcome. The unconfoundedness assumption makes sure that we observe all hidden confounders

and hence are able to adjust for them when estimating individualised treatment effects. This is difficult, as having no hidden confounders is a strong assumption to make and cannot be verified in practice. It is crucial to note that confounders; and by extension hidden confounders; affect all modelling in data science, and this problem is by no means unique to causal inference. Without adjusting for hidden confounders in electronic health records, standard methods would otherwise spuriously identify non-causal treatments. Because counterfactuals are never observed, it is not possible to test for the existence of hidden confounders that might affect them. Several methods have been proposed to address this, to either measure the potential effect that an unmeasured confounder could have, or to make use of proxy variables to infer latent variables that can act as substitutes for hidden confounders, and consideration will be given to both.

Sensitivity analyses can be used to evaluate the potential impact an unmeasured confounder could have on the estimation of treatment effects, and as such this stage would happen after the model is created. We will look at the work of Franks et al [58]. We will factorize the joint distribution of observed and unobserved patient outcomes into separate factors that are nonparametrically identified from the data. This enables quantifying the dependence among the partially observable patient outcomes, treatments, and patient features. In this sense, we can determine the suitability of using treatment effects methods for both MIMIC and the Systemwide dataset.

We will also consider the work of Wang and Blei [59]. We will use their "deconfounder", an algorithm that combines unsupervised machnine learning and predictive model checking to perform causal inference in multipl-cause settings. We will infer a latent variable as a substitute for unobserved confounders and then use that substitute to perform causal inference. It requires weaker assumptions than classical causal inference. What can also be done is to use domain knowledge of clinicians to assess the validity of this assumption, as well as using their prior knowledge of causal links among patient characteristics, and this is discussed in the co-design section that follows.

5.8 Co-design

The project will work heavily with the existing HDR-UK South West Better Care Partnership. Part of this project involves ongoing co-creation workshops; events with clinicians and stakeholders. These events will provide opportunity to assess user requirements for a system such as the one proposed above, such that the model developed is clinically interpretable, and appropriate for clinical validation and implementation. These events will take place every 4 months during the life cycle of the project.

5.8.1 Data-specific queries

Work to date with the LABMARCS study has made extensive use of the clinicians' insight. I arranged a series of sessions with clinicians to discuss peculiarities in the data, discuss existing treatment protocols, and to inform feature selection. When considering missing values and data imputation, many algorithms make the assumption that if data is missing, it is missing at random. In medicine, missingness is often informative. Quite often a test is only conducted if a condition is suspected, and so the measurement value is correlated with its presence or absence. Imputing this might give a spurious value. By talking with clinicians, we were able to narrow down a list of laboratory measurements that were taken on a regular basis. We could then make the assumption that these values were missing at random and therefore include them in our imputation.

There will be a few sessions over 2 months as the MIMIC dataset is explored where there will be opportunities to query features of the data that will inform the design of the clinical decision support system, and this process will be repeated as and when the Systemwide dataset becomes available.

5.8.2 Prior knowledge of causal links

As mentioned earlier when testing key assumptions, prior knowledge of causal links among patient characteristics, treatments and outcomes will be leveraged to identify potential unobserved confounders. Clinicians and other stakeholders involved in the management of antimicrobials will be consulted to check if all confounding variables are included in the observational data.

After data-specific queries, co-design sessions will focus on causal links in the static setting, in preparation for the first problem of optimal antimicrobial prescribing in the static setting. After this we will start to consider time-confounders and causal links in the temporal setting in the lead up to the second problem when we not only consider the optimal antimicrobial but the timing of this.

5.9 Causal discovery

All that is being done here is a systematic approach to the problem, as well as presentation of the causal relationships (inclusive of confounders and mediators) in a diagrammatic; and therefore more understandable; form. We will make use of the recommendations of Tennant et al [60] in the use of directed acyclic graphs in applied health research to present diagrammatic representations of the causal relationships amongst features to both aid clinicians in the causal discovery process, as well as in published work to allow for greater scrutiny.

5.10 Missing data

A substantial challenge with the data being used is the large amount of missing data, particularly in the parameters of interest in the study. The limited sample size and the high proportion of missing data would make it unfeasible to exclude those patients with missing data. Methods used in many prediction models rely heavily upon imputation. However, the form of imputation used relies heavily upon the type of missingness. Many forms of imputation used previously assume the data is missing at random or missing completely at random, such as feedforward, mean imputation, and nearest neighbour methods. There is considerable evidence that many clinical data are missing not at random, with examples including the administration of antibiotics when an infection is suspected, or frequency of observations being increased when a patient's condition deteriorates. Most forms of imputation are not suitable for data missing not at random [61]. Informative missingness in electronic health records can also be a problem. Data missing not at random is informative, and this informativeness can be incorporated into the clinical prediction model. However, the transportability of the missing data mechanism is a concern, as this could be compromised once the model is deployed and the predictive value of certain variables becomes known

[62]. Che et. al. propose the use of recurrent neural networks for multivariate time series with missing data [63]. Whilst their work incorporated the missing values as predictors, they did not use the model explicitly for imputation. My intention is to investigate this form of model for imputation and to evaluate it's effectiveness.

Work with the LABMARCS project involved assessment of missingness as well as determining

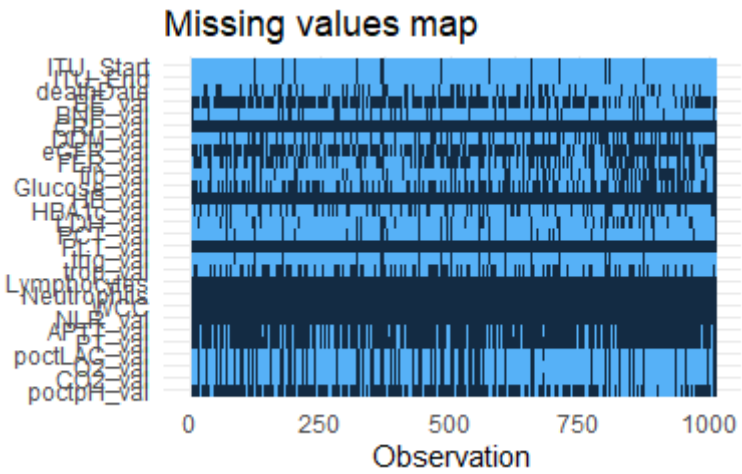


Figure 15: Missingness for all patients and variables, where missingness is absence of a variable at any point of a patient's stay

what values were suitable for imputation. Many values were determined to be missing not at random, as they were often ordered when the condition the test confirmed was suspected. This would be an example of assignment bias. These variables were not considered for imputation. Data was combined into both the long and wide formats for suitability for the different imputation types, and both MICE and MRNN were performed.

MRNN (Multi-directional Recurrent Neural networks) are a very new and exciting form of imputation that estimate missing data on temporal data streams [64].They can handle different and irregular times, something common in medicine. Most imputation approaches either interpolate within data streams or impute across data streams. This method does both simultaneously and so doesn't miss important data.

In the context of LABMARCS, there were many considerations for the suitability of MRNN, including:

- Sample size (MRNN performs poorly at low sample sizes, below 7000) Note this is specific to MIMIC, and MRNN outperforms other algorithms in other small sample datasets)
- Frequency of measurement (fewer measurements per patient degrades performance of MRNN)
- Dimensionaltiy of the data
- Number of measurements per patient
- Amount of missing data

MRNN gave limited improvements in performance at low missingness, but marked improvement in performance when missingness was high, relative to other imputation methods. In our case missingness was high. We also had a low number of measurements per patient, but this was still high enough to outperform other techniques.

5.11 Incremental approach

In the diagram below I attempt to comprehensively address all the possible questions that could be asked regarding antimicrobial treatment choice. Not all of these will be able to be addressed during this PhD. Complexity comes in a few forms:

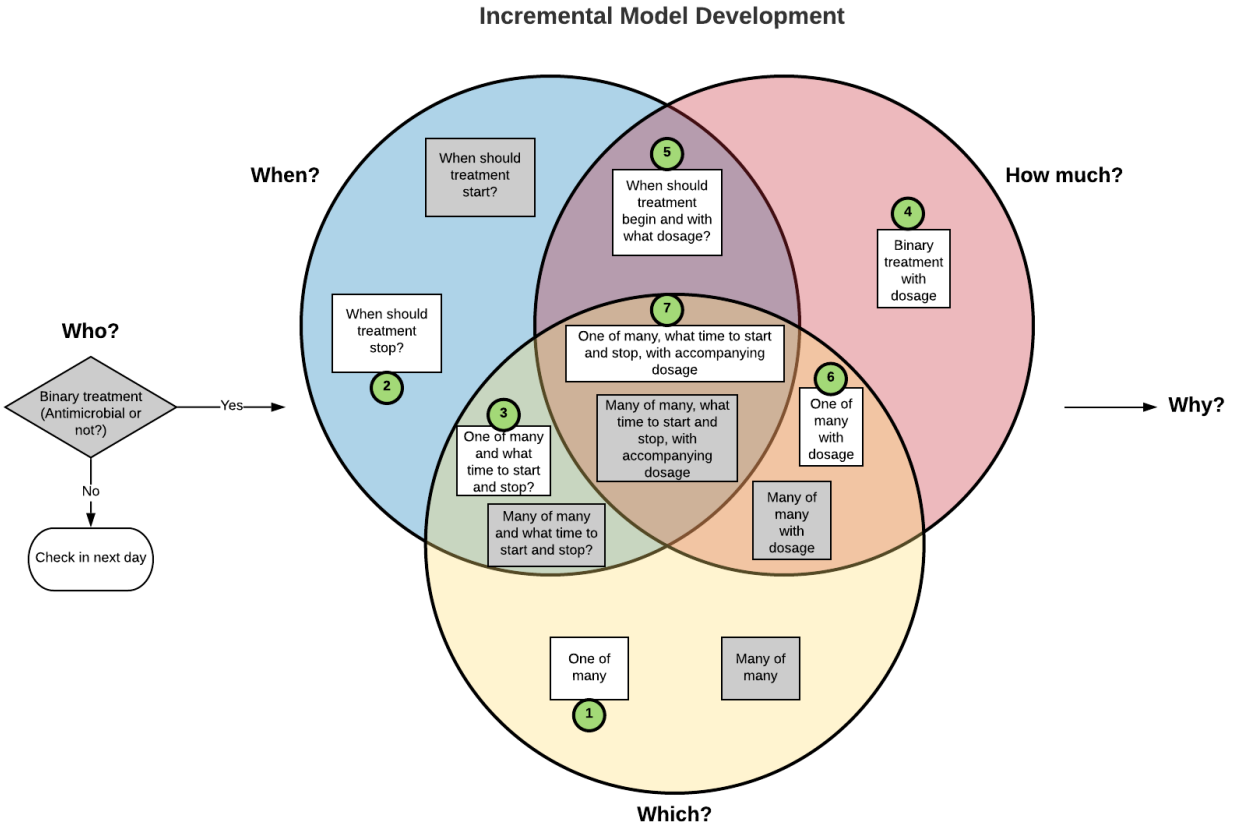


Figure 16: Grey boxes are questions that are outside the scope of this project. White boxes are questions that are addressed during this project. Finally, the numbers correspond to the order in which these questions are addressed in this project

- Static to temporal setting
- From binary treatment, to one of many, to multiple treatment options
- Consideration of dosage

The idea is to address these questions from simplest to most complex, and the 3 main problems I have decided to address follow this incremental approach and group some of these questions. Finally, the door is left open for extensions, covering questions that may not be addressed during the course of this PhD.

5.12 Problem 1: Optimal antimicrobial in cross-sectional setting

In the first problem we will address just one question: "What is the optimal antimicrobial choice in the static (cross-sectional) setting?" There are a number of considerations to take into account when deciding what model is appropriate to answer this question, and the vast majority of these are also relevant to the other problems. Notably, how they model the treatments in terms of response surfaces, as well as how they handle selection bias.

We will estimate treatment response surfaces in a supervised learning setting by fitting two regression models for treated and untreated patients, with a shared structure between the two response functions and separate outcome models for different treatments.

In terms of handling the treatment selection bias, we will build "balancing representations" or treatment invariant representations, which are representations of the patient covariates that

have equal probabilities under the possible treatment options. Conditioning on the balancing representations to estimate the patient outcomes reduces the selection bias present.

Approximately 4-5 months will be set aside to solve this research problem, excluding paper write-up time.

5.12.1 Small sample regime

With respect to Systemwide dataset, which will have a smaller sample than MIMIC, it is vital to handle selection bias, and to share training data between response surfaces, by either fitting a regression model for the two response surfaces with the treatment as an input feature, or by using multi-task learning to estimate the response surfaces.

We will consider the use of CMGP, NSGP, Counterfactual Regression, and Similarity Preserved Individual Treatment Effect in this setting. CMGP and NSGP have the benefit of confidence intervals, which is important for our use case.

5.12.2 Large sample regime

When the observational data has enough samples, selection bias no longer plays a more crucial role: the way in which outcomes are modeled and how hyperparameter tuning is performed. So, using outcome-specific hyperparameters becomes more important.

We will consider the use of Treatment-Agnostic Representation Network and GANITE in this setting.

5.13 Problem 2: Optimal antimicrobial in temporal setting (timing)

In the second problem we will address two questions:

- What is the optimal time to stop a single antimicrobial treatment?
- What is the optimal choice of antimicrobial at initiation of treatment?
- What is the optimal choice of antimicrobial at each time step?

We will not be addressing the ideal time to start treatment, as we showed earlier that faster administration of antimicrobials leads to better patient outcomes. Therefore there is nothing to be gained from asking this question. The speed at which antimicrobials are administered depends primarily on early recognition of sepsis, and this is outside the scope of this project. The PNEWS score will be used to help identify those at risk of sepsis earlier, and this will certainly be utilised in this project in terms of the ground truth of who does/doesn't have sepsis and by extension those that then require antimicrobials.

5.13.1 Time-dependent confounders

There is also time-dependent confounding bias in this setting. That is to say patient covariates that are affected by past treatments and which then influence future treatments and observations. In this case, overlap means that at each timestep we assume that each treatment option has a non-zero probability of being administered. The overlap assumption can be assessed by computing the probability of a patient for receiving each possible treatment at each time point. Moreover, the unconfoundedness assumption in the sequential setting means that at each timestep we observe all variables affecting the patient's treatment and outcome. Validating this assumption is the same as in the static setting.

5.13.2 Model choice

Model choice in this setting depends on the method of handling time-dependent confounding bias, handling discrete-time or continuous-time interventions, and modelling the progression of one or more patient covariates of interest.

We will be considering the use of Counterfactual Recurrent Networks. They use balancing representations to handle time-dependent confounders, and handle time in the discrete setting, which matches up with daily ward rounds, which is the point at which treatment decisions are made regarding changes in antimicrobial therapy. CRN estimates individualized treatment effects in the discrete-time setting and allows for flexible modeling of the patient's baseline information and progression of multiple clinical variables and treatments of interest.

Approximately 4-5 months will be set aside to solve this research problem, excluding paper write-up time.

5.14 Problem 3: Optimal dosing

Finallly, in the third problem, we will look at the following questions:

- What is the optimal dose for a single antimicrobial in the static setting?
- What is the optimal dose for a single antimicrobial at each time step?
- What is the optimal dose and antimicrobial at initiation of treatment?
- What is the optimal dose and antimicrobial at each time step?

We will use machine learning methods to estimate individualised dose-response curves. Estimating these curves allows us to determine the best treatment-dosage pair for each patient. This is useful in our setting as the relationship amongst antimicrobial toxicity, efficacy and heterogenous patient features are complex.

Because there are a potentially infinite number of outcomes from dosage (continuously valued), it's not possible to estimate a response surface for each possible treatment-dosage outcome. We must also consider both treatment assignment bias and dosage assignment bias.

We will look at using SCIGAN, a model that uses generative adversarial networks to learn the data distribution of the counterfactual outcomes and thus generate individualised dose-response curves. It doesn't place any restrictions on the form of the treatment-dose response function. It has previously been used to consider the effect of antibiotic dosage on patient outcomes, but not in the context of sepsis.

Approximately 4-5 months will be set aside to solve this research problem, excluding paper write-up time.

5.15 Evaluation

In standard supervised learning one computes the prediction error on a validation dataset, and then selects the model with the lowest error. However, because counterfactuals are not observed, it is not possible to compute error on estimating the ground truth causal effects on real data.

5.15.1 Synthetic data

Synthetic data is an initial step for evaluating our machine learning model. Semi-synthetic datasets with real patient features and simulated outcomes will be created. The outcomes need to be simulated in order to have access to the counterfactuals. To create a synthetic observational dataset, parametric functions such as polynomials or exponentials will be used to generate patient outcomes under different treatment options. Confounding bias is introduced by assigning treatments according to Bernouilli or Categorical random variable, with parameters depending on patient characteristics. Then, the causal inference methods can be evaluated based on their error on estimating all potential outcomes and the ground truth causal effects, where ground truth causal effects are the difference between the patient outcome when receiving treatment and the patient outcome when not receiving treatment. In real data, it is only possible to observe one of these outcomes. The sensitivity of the causal inference methods to confounding bias can be assessed by changing the parameters of the probability distributions used to assign treatments for the patients in the observational data generated for training. The purpose of these initial steps is to have a preliminary evaluation to assess the robustness of a suite of potential models at estimating individualised treatment effects, but it does not capture the complexities of real-world data. Guidelines by Alaa et al show how validation through synthetic data allows us to narrow down the choice of suitable causal inference methods for the observational data available [48]. However, we need data-driven validation procedures similar to cross-validation in traditional machine learning research to find the most suitable model for a real-world study.

5.15.2 Internal model validation

The proposed method for assessing causal inference models will be with influence functions; a technique in robust statistics and efficiency theory. This estimates the loss of machine learning models for causal inference without requiring counterfactual data. Model selection is an important first step before this. Influence functions approximate the error of causal inference methods on **held-out test dataset** and subsequently enables us to select the best performing model for an observational dataset of interest.

5.15.3 Expert validation

Real world data pertains to the routinely collected data relating to a patient's health status. The analysis of these data regarding usage and effectiveness is known as real world evidence. Work by Beaulieu-Jones et al outlines the use of such data, and emphasises the importance of expert knowledge with respect to the validation of a model's estimate of treatment effects [65]. During one of the co-design sessions, a "dummy ward round" will be conducted. A series of recommendations; along with any accompanying explanations; will be presented to clinicians and antimicrobial stewardship professionals involved, and they will specify if the decisions are clinically sound. More work will be done to establish the format of this.

5.15.4 External model validation

Whilst this is not in the scope of this PhD, the design of the methodology and the project as a whole has this in mind. The purpose of internal validation is to provide a robust chain of reasoning that this model may be effective in practice. The machine learning model for causal inference would be evaluated in real clinical practice by measuring whether the patient outcomes have been improved when assigning treatments according to the recommendation of the causal inference model. One could imagine a dashboard of the recommendations of the clinical decision support system. Treatment recommendations would be provided and the doctor

would specify whether or not they accepted the recommendation. One could then not only look at the outcomes before and after the implementation of the system, but also outcomes when acting in line with recommendations. This follows the "doctor in the loop" paradigm.

5.15.5 Deployment / Implementation

Factors associated with adoption of a CDSS such as clinicians' willingness, system uptake, organisational limitations and implementation constraints must be considered. Clinician compliance and uptake is a major concern for the adoption of health-related technologies and can cause interventions to fail to provide benefits in the long term [66]. A common criticism of prediction models with sepsis is their lack of investigation of direct clinical action associated with the sepsis prediction [67].

5.15.6 Output / visualisation

Performance Predictions Uncertainties Interpretations

Performance metrics:

Bedside prediction models are better judged by properly contextualised positive and negative predictive values. Priorities for prediction are context specific but involves tradeoffs determined by consequences of: • False positives - Overtreatment and poor antimicrobial stewardship • False negatives - Missed cases It is important that clinicians uptake this sort of technology, and so the number of false negatives and false positives will be specified by clinicians such that the requirements are clinically meaningful, for example "no more than X number of missed cases per day"

5.16 Extensions

5.16.1 RCT data

5.16.2 Pharmacometric and QSP approaches

5.17 Responsible innovation

5.17.1 Research governance

This study will be conducted in accordance with: • TRIPOD Guidelines - Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis • Joanna Briggs Institute Checklist for Analytical Cross Sectional Studies • Grading of Recommendations Assessment, Development and Evaluation • The international conference for harmonisation of good clinical practice (ICH GCP) • The research governance framework for health and social care • The declaration of Helsinki • Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals41

5.17.2 Ethics

This study will undergo rigorous ethical and regulatory approval process and a favourable opinion will be sought from an NHS REC (UK Health Departments Research Ethics Service NHS REC). The laboratory tests are performed as part of routine clinical care. No extra tests will be performed on existing samples. Therefore, we do not anticipate any major ethical issues.

Systemwide dataset request flowchart? MIMIC approval

5.17.3 Assessment and management of risk

Risk	Description	Mitigation
1	Predictions not precise and accurate enough for clinical use	At a population level we have a large amount of data for learning so imprecise predictions are unlikely. For small or unusual patient populations predictions may be imprecise or biased, hence need for clinical validation on external datasets
2	Issues with data access & Data governance	Arrangements are already in place, regular meetings arranged to ensure issues are dealt with regularly.
3	Difficulties in linking data from primary and secondary care	BNSSG dataset already exists, together with pathology dataset.

5.17.4 Open Science

It is our intention to share reusable code and any platforms developed via GitHub.

5.18 Public engagement

The Centre of Doctoral Training for Digital Health & Care has an ongoing residency at "We the Curious", a science museum and interactive space. I developed an activity here in August 2021 where we designed interactive activities that helped children and adults engage with data science ethics. We had post it notes where the public would give their thoughts on how comfortable they would feel with their data being used for research, how they would feel being treated by a "Robo-Doctor" and other thought experiments. This happens for a week every year, and more project-specific activities will be developed in the summer of '22 and '23.

5.19 National importance

There is considerable support for the digitisation of the NHS, with the Wachter report making recommendations for all trusts to be digitised by 2023 [68], and the NHS Long Term Plan intending to go fully paperless by 2024 [69]. There is also a real push towards the adoption of artificial intelligence in the NHS, including the use of clinical decision support systems, with £100 million having been invested in the NHSX's AI Lab. This project falls under the remit of the EPSRC's research portfolio, Clinical Technologies (excluding imaging). It is the intention to submit to journals both in the medical area and computer science, to maximise the impact of the research. Digital health is an emerging area, and is very multidisciplinary. It is key that the clinicians and other key stakeholders such as commissioners will have access to the kind of research that will benefit them, as well as computer scientists and other academics who may be able to build upon the proposed research. This research will take place in the context of an internationally competitive area. Should the work be successful, it is the intention that any models be able to scale, and we will seek to collaborate with other organisations to validate data externally.

6 Bibliography

References

[1] R. Phillip Dellinger, Mitchell M. Levy, Andrew Rhodes, Djillali Annane, Herwig Gerlach, Steven M. Opal, Jonathan E. Sevransky, Charles L. Sprung, Ivor S. Douglas, Roman Jaeschke, Tiffany M. Osborn, Mark E. Nunnally, Sean R. Townsend, Konrad Reinhart, Ruth M. Kleinpell, Derek C. Angus, Clifford S. Deutschman, Flavia R. Machado, Gordon D. Rubenfeld, Steven A. Webb, Richard J. Beale, Jean-Louis Vincent, Rui Moreno, and and the Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. "Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2012". en-US. In: *Critical Care Medicine* 41.2 (Feb. 2013), pp. 580–637. ISSN: 0090-3493. DOI: 10.1097/CCM.0b013e31827e83af. URL: https://journals.lww.com/ccmjournal/Fulltext/2013/02000/Surviving_Sepsis_Campaign__International.24.aspx (visited on 11/02/2020) (page 2).

[2] Kristina E. Rudd, Sarah Charlotte Johnson, Kareha M. Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V. Colombara, Kevin S. Ikuta, Niranjan Kissoon, Simon Finfer, Carolin Fleischmann-Struzek, Flavia R. Machado, Konrad K. Reinhart, Kathryn Rowan, Christopher W. Seymour, R. Scott Watson, T. Eoin West, Fatima Marinho, Simon I. Hay, Rafael Lozano, Alan D. Lopez, Derek C. Angus, Christopher J. L. Murray, and Mohsen Naghavi. "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study". English. In: *The Lancet* 395.10219 (Jan. 2020). Publisher: Elsevier, pp. 200–211. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(19)32989-7. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32989-7/abstract (visited on 10/26/2020) (page 2).

[3] Talha Khan Burki. "Sharp rise in sepsis deaths in the UK". English. In: *The Lancet Respiratory Medicine* 6.11 (Nov. 2018). Publisher: Elsevier, p. 826. ISSN: 2213-2600, 2213-2619. DOI: 10.1016/S2213-2600(18)30382-5. URL: https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(18)30382-5/abstract (visited on 11/02/2020) (page 2).

[4] "Antimicrobial resistance". en. URL: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance (visited on 11/07/2020) (page 3).

[5] Marlieke E. A. de Kraker, Andrew J. Stewardson, and Stephan Harbarth. "Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?" In: *PLoS Medicine* 13.11 (Nov. 2016). ISSN: 1549-1277. DOI: 10.1371/journal.pmed.1002184. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5127510/ (visited on 10/27/2020) (page 3).

[6] "English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) report". en. URL: https://www.gov.uk/government/publications/english-surveillance-programme-antimicrobial-utilisation-and-resistance-espaur-report (visited on 11/03/2020) (page 3).

[7] Katherine K. Perez, Randall J. Olsen, William L. Musick, Patricia L. Cernoch, James R. Davis, Geoffrey A. Land, Leif E. Peterson, and James M. Musser. "Integrating rapid pathogen identification and antimicrobial stewardship significantly decreases hospital costs". eng. In: *Archives of Pathology & Laboratory Medicine* 137.9 (Sept. 2013), pp. 1247–1254. ISSN: 1543-2165. DOI: 10.5858/arpa.2012-0651-OA (page 3).

[8] Andrew Rhodes, Laura E. Evans, Waleed Alhazzani, Mitchell M. Levy, Massimo Antonelli, Ricard Ferrer, Anand Kumar, Jonathan E. Sevransky, Charles L. Sprung, Mark E. Nunnally, Bram Rochwerg, Gordon D. Rubenfeld, Derek C. Angus, Djillali Annane, Richard J. Beale, Geoffrey J. Bellinghan, Gordon R. Bernard, Jean-Daniel Chiche, Craig Coopersmith, Daniel P. De Backer, Craig J. French, Seitaro Fujishima, Herwig Gerlach, Jorge Luis Hidalgo, Steven M. Hollenberg, Alan E. Jones, Dilip R. Karnad, Ruth M. Kleinpell, Younsuck Koh, Thiago Costa Lisboa, Flavia R. Machado, John J. Marini, John C. Marshall, John E. Mazuski, Lauralyn A. McIntyre, Anthony S. McLean, Sangeeta Mehta, Rui P. Moreno, John Myburgh, Paolo Navalesi, Osamu Nishida, Tiffany M. Osborn, Anders Perner, Colleen M. Plunkett, Marco Ranieri, Christa A. Schorr, Maureen A. Seckel, Christopher W. Seymour, Lisa Shieh, Khalid A. Shukri, Steven Q. Simpson, Mervyn Singer, B. Taylor Thompson, Sean R. Townsend, Thomas Van der Poll, Jean-Louis Vincent, W. Joost Wiersinga, Janice L. Zimmerman, and R. Phillip Dellinger. "Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016". en-US. In: *Critical Care Medicine* 45.3 (Mar. 2017), pp. 486–552. ISSN: 0090-3493. DOI: 10.1097/CCM.0000000000002255. URL: https://journals.lww.com/ccmjournal/Fulltext/2017/03000/Surviving_Sepsis_Campaign___International.15.aspx (visited on 07/07/2020) (pages 3, 19).

[9] Anand Kumar, Daniel Roberts, Kenneth E. Wood, Bruce Light, Joseph E. Parrillo, Satendra Sharma, Robert Suppes, Daniel Feinstein, Sergio Zanotti, Leo Taiberg, David Gurka, Aseem Kumar, and Mary Cheang. "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock". eng. In: *Critical Care Medicine* 34.6 (June 2006), pp. 1589–1596. ISSN: 0090-3493. DOI: 10.1097/01.CCM.0000217961.75225.E9 (page 3).

[10] L. Leibovici, M. Drucker, H. Konigsberger, Z. Samra, S. Harrari, S. Ashkenazi, and S. D. Pitlik. "Septic shock in bacteremic patients: risk factors, features and prognosis". eng. In: *Scandinavian Journal of Infectious Diseases* 29.1 (1997), pp. 71–75. ISSN: 0036-5548. DOI: 10.3109/00365549709008668 (page 3).

[11] Emanuel Rivers, Bryant Nguyen, Suzanne Havstad, Julie Ressler, Alexandria Muzzin, Bernhard Knoblich, Edward Peterson, and Michael Tomlanovich. "Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock". In: *New England Journal of Medicine* 345.19 (Nov. 2001). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa010307, pp. 1368–1377. ISSN: 0028-4793. DOI: 10.1056/NEJMoa010307. URL: https://doi.org/10.1056/NEJMoa010307 (visited on 07/07/2020) (page 3).

[12] R. Menéndez, A. Torres, S. Reyes, R. Zalacain, A. Capelastegui, J. Aspa, L. Borderías, J. J. Martín-Villasclaras, S. Bello, I. Alfageme, F. R. de Castro, J. Rello, L. Molinos, and J. Ruiz-Manzano. "Initial management of pneumonia and sepsis: factors associated with improved outcome". eng. In: *The European Respiratory Journal* 39.1 (Jan. 2012), pp. 156–162. ISSN: 1399-3003. DOI: 10.1183/09031936.00188710 (page 3).

[13] Maureen Campion and Gail Scully. "Antibiotic Use in the Intensive Care Unit: Optimization and De-Escalation". en. In: *Journal of Intensive Care Medicine* 33.12 (Dec. 2018). Publisher: SAGE Publications Inc STM, pp. 647–655. ISSN: 0885-0666. DOI: 10.1177/0885066618762747. URL: https://doi.org/10.1177/0885066618762747 (visited on 10/06/2021) (page 4).

[14] Vincent X. Liu, Vikram Fielding-Singh, John D. Greene, Jennifer M. Baker, Theodore J. Iwashyna, Jay Bhattacharya, and Gabriel J. Escobar. "The Timing of Early Antibiotics and Hospital Mortality in Sepsis". eng. In: *American Journal of Respiratory and Critical*

Care Medicine 196.7 (2017), pp. 856–863. ISSN: 1535-4970. DOI: 10.1164/rccm.201609-1848OC (page 4).

[15] Aaron Sango, Yvette S. McCarter, Donald Johnson, Jason Ferreira, Nilmarie Guzman, and Christopher A. Jankowski. "Stewardship approach for optimizing antimicrobial therapy through use of a rapid microarray assay on blood cultures positive for Enterococcus species". eng. In: *Journal of Clinical Microbiology* 51.12 (Dec. 2013), pp. 4008–4011. ISSN: 1098-660X. DOI: 10.1128/JCM.01951-13 (page 5).

[16] Nick Daneman, Asgar H. Rishu, Wei Xiong, Sean M. Bagshaw, Deborah J. Cook, Peter Dodek, Richard Hall, Anand Kumar, Francois Lamontagne, Francois Lauzier, John C. Marshall, Claudio M. Martin, Lauralyn McIntyre, John Muscedere, Steven Reynolds, Henry T. Stelfox, Robert A. Fowler, and Canadian Critical Care Trials Group. "Bacteremia Antibiotic Length Actually Needed for Clinical Effectiveness (BALANCE): study protocol for a pilot randomized controlled trial". eng. In: *Trials* 16 (Apr. 2015), p. 173. ISSN: 1745-6215. DOI: 10.1186/s13063-015-0688-z (page 5).

[17] Juan Pasquau. "The Reduction in Duration of Antibiotic Therapy as a Key Element of Antibiotic Stewardship Programs". In: *Journal of Antimicrobial Agents* 01 (Jan. 2015). DOI: 10.4172/2472-1212.1000103 (page 5).

[18] Thomas C. Havey, Robert A. Fowler, and Nick Daneman. "Duration of antibiotic therapy for bacteremia: a systematic review and meta-analysis". eng. In: *Critical Care (London, England)* 15.6 (2011), R267. ISSN: 1466-609X. DOI: 10.1186/cc10545 (page 5).

[19] Rachida el Moussaoui, Corianne A. J. M. de Borgie, Peterhans van den Broek, Willem N. Hustinx, Paul Bresser, Guido E. L. van den Berk, Jan-Werner Poley, Bob van den Berg, Frans H. Krouwels, Marc J. M. Bonten, Carla Weenink, Patrick M. M. Bossuyt, Peter Speelman, Brent C. Opmeer, and Jan M. Prins. "Effectiveness of discontinuing antibiotic treatment after three days versus eight days in mild to moderate-severe community acquired pneumonia: randomised, double blind study". eng. In: *BMJ (Clinical research ed.)* 332.7554 (June 2006), p. 1355. ISSN: 1756-1833. DOI: 10.1136/bmj.332.7554.1355 (page 5).

[20] Yoshiro Hayashi and David L. Paterson. "Strategies for reduction in duration of antibiotic use in hospitalized patients". eng. In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 52.10 (May 2011), pp. 1232–1240. ISSN: 1537-6591. DOI: 10.1093/cid/cir063 (page 5).

[21] Scott T. Micek, Emily C. Welch, Junaid Khan, Mubashir Pervez, Joshua A. Doherty, Richard M. Reichley, and Marin H. Kollef. "Empiric combination antibiotic therapy is associated with improved outcome against sepsis due to Gram-negative bacteria: a retrospective analysis". eng. In: *Antimicrobial Agents and Chemotherapy* 54.5 (May 2010), pp. 1742–1748. ISSN: 1098-6596. DOI: 10.1128/AAC.01365-09 (page 6).

[22] Nicola Petrosillo, Alessandro Capone, Stefano Di Bella, and Fabrizio Taglietti. "Management of antibiotic resistance in the intensive care unit setting". eng. In: *Expert Review of Anti-Infective Therapy* 8.3 (Mar. 2010), pp. 289–302. ISSN: 1744-8336. DOI: 10.1586/eri.10.7 (page 6).

[23] Peter M. Rothwell. "External validity of randomised controlled trials: "To whom do the results of this trial apply?"" English. In: *The Lancet* 365.9453 (Jan. 2005). Publisher: Elsevier, pp. 82–93. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(04)17670-8. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(04)17670-8/fulltext (visited on 10/11/2021) (page 9).

[24] Laura E. Bothwell, Jeremy A. Greene, Scott H. Podolsky, and David S. Jones. "Assessing the Gold Standard–Lessons from the History of RCTs". eng. In: *The New England Journal of Medicine* 374.22 (June 2016), pp. 2175–2181. ISSN: 1533-4406. DOI: 10.1056/NEJMms1604593 (page 9).

[25] Thomas R. Frieden. "Evidence for Health Decision Making — Beyond Randomized, Controlled Trials". In: *New England Journal of Medicine* 377.5 (Aug. 2017). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMra1614394, pp. 465–475. ISSN: 0028-4793. DOI: 10.1056/NEJMra1614394. URL: https://doi.org/10.1056/NEJMra1614394 (visited on 10/11/2021) (page 9).

[26] L. F. Hutchins, J. M. Unger, J. J. Crowley, C. A. Coltman, and K. S. Albain. "Underrepresentation of patients 65 years of age or older in cancer-treatment trials". eng. In: *The New England Journal of Medicine* 341.27 (Dec. 1999), pp. 2061–2067. ISSN: 0028-4793. DOI: 10.1056/NEJM199912303412706 (page 9).

[27] A. Ramamoorthy, M. A. Pacanowski, J. Bull, and L. Zhang. "Racial/ethnic differences in drug disposition and response: review of recently approved drugs". eng. In: *Clinical Pharmacology and Therapeutics* 97.3 (Mar. 2015), pp. 263–273. ISSN: 1532-6535. DOI: 10.1002/cpt.61 (page 9).

[28] A. M. Alaa, J. Yoon, S. Hu, and M. van der Schaar. "Personalized Risk Scoring for Critical Care Prognosis Using Mixtures of Gaussian Processes". In: *IEEE Transactions on Biomedical Engineering* 65.1 (Jan. 2018). Conference Name: IEEE Transactions on Biomedical Engineering, pp. 207–218. ISSN: 1558-2531. DOI: 10.1109/TBME.2017.2698602 (page 9).

[29] Neelja Singhal, Manish Kumar, Pawan K. Kanaujia, and Jugsharan S. Virdi. "MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis". In: *Frontiers in Microbiology* 6 (Aug. 2015). ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00791. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525378/ (visited on 11/07/2020) (page 9).

[30] Georgia Vrioni, Constantinos Tsiamis, George Oikonomidis, Kalliopi Theodoridou, Violeta Kapsimali, and Athanasios Tsakris. "MALDI-TOF mass spectrometry technology for detecting biomarkers of antimicrobial resistance: current achievements and future perspectives". In: *Annals of Translational Medicine* 6.12 (June 2018). ISSN: 2305-5839. DOI: 10.21037/atm.2018.06.28. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6046294/ (visited on 11/07/2020) (page 9).

[31] Tsung-Yun Hou, Chuan Chiang-Ni, and Shih-Hua Teng. "Current status of MALDI-TOF mass spectrometry in clinical microbiology". en. In: *Journal of Food and Drug Analysis*. Mass Spectrometry for Clinical Diagnosis 27.2 (Apr. 2019), pp. 404–414. ISSN: 1021-9498. DOI: 10.1016/j.jfda.2019.01.001. URL: http://www.sciencedirect.com/science/article/pii/S1021949819300146 (visited on 11/07/2020) (page 9).

[32] Mah Laka, Adriana Milazzo, and Tracy Merlin. "Can evidence-based decision support tools transform antibiotic management? A systematic review and meta-analyses". en. In: *Journal of Antimicrobial Chemotherapy* 75.5 (May 2020). Publisher: Oxford Academic, pp. 1099–1111. ISSN: 0305-7453. DOI: 10.1093/jac/dkz543. URL: https://academic.oup.com/jac/article/75/5/1099/5710709 (visited on 10/26/2020) (page 9).

[33] Michael A. Rubin, Kim Bateman, Sharon Donnelly, Gregory J. Stoddard, Kurt Stevenson, Reed M. Gardner, and Matthew H. Samore. "Use of a personal digital assistant for managing antibiotic prescribing for outpatient respiratory tract infections in rural com-

munities". eng. In: *Journal of the American Medical Informatics Association: JAMIA* 13.6 (Dec. 2006), pp. 627–634. ISSN: 1067-5027. DOI: 10.1197/jamia.M2029 (page 10).

[34] Matthew H. Samore, Kim Bateman, Stephen C. Alder, Elizabeth Hannah, Sharon Donnelly, Gregory J. Stoddard, Bassam Haddadin, Michael A. Rubin, Jacquelyn Williamson, Barry Stults, Randall Rupper, and Kurt Stevenson. "Clinical decision support and appropriateness of antimicrobial prescribing: a randomized trial". eng. In: *JAMA* 294.18 (Nov. 2005), pp. 2305–2314. ISSN: 1538-3598. DOI: 10.1001/jama.294.18.2305 (page 10).

[35] "A critical assessment of published guidelines and other decision-support systems for the antibiotic treatment of community-acquired respiratory tract infections - Clinical Microbiology and Infection". URL: https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(15)30085-9/fulltext (visited on 11/08/2020) (page 10).

[36] Lothar Gierl, Dagmar Steffen, Dusan Ihracky, and Rainer Schmidt. "Methods, architecture, evaluation and usability of a case-based antibiotics advisor". en. In: *Computer Methods and Programs in Biomedicine* 72.2 (Oct. 2003), pp. 139–154. ISSN: 0169-2607. DOI: 10.1016/S0169-2607(02)00121-9. URL: http://www.sciencedirect.com/science/article/pii/S0169260702001219 (visited on 11/08/2020) (page 10).

[37] Victor L. Yu, S. M. Miller, M. M. Beattie, and A. A. Butt. "Personal Digital Assistant Infectious Diseases Applications for Health Care Professionals". en. In: *Clinical Infectious Diseases* 36.8 (Apr. 2003). Publisher: Oxford Academic, pp. 1018–1029. ISSN: 1058-4838. DOI: 10.1086/368198. URL: https://academic.oup.com/cid/article/36/8/1018/446938 (visited on 11/08/2020) (page 10).

[38] Mical Paul, Steen Andreassen, Evelina Tacconelli, Anders D. Nielsen, Nadja Almanasreh, Uwe Frank, Roberto Cauda, Leonard Leibovici, and on behalf of the TREAT Study Group. "Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial". en. In: *Journal of Antimicrobial Chemotherapy* 58.6 (Dec. 2006). Publisher: Oxford Academic, pp. 1238–1245. ISSN: 0305-7453. DOI: 10.1093/jac/dkl372. URL: https://academic.oup.com/jac/article/58/6/1238/728601 (visited on 10/26/2020) (page 10).

[39] Zafer Tandogdu, Evgenios T. A. Kakariadis, Kurt Naber, Florian Wagenlehner, and Truls Erik Bjerklund Johansen. "Appropriate empiric antibiotic choices in health care associated urinary tract infections in urology departments in Europe from 2006 to 2015: A Bayesian analytical approach applied in a surveillance study". en. In: *PLOS ONE* 14.4 (Apr. 2019). Publisher: Public Library of Science, e0214710. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0214710. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214710 (visited on 10/26/2020) (page 10).

[40] Idan Yelin, Olga Snitser, Gal Novich, Rachel Katz, Ofir Tal, Miriam Parizade, Gabriel Chodick, Gideon Koren, Varda Shalev, and Roy Kishony. "Personal clinical history predicts antibiotic resistance of urinary tract infections". en. In: *Nature Medicine* 25.7 (July 2019). Number: 7 Publisher: Nature Publishing Group, pp. 1143–1152. ISSN: 1546-170X. DOI: 10.1038/s41591-019-0503-6. URL: https://www.nature.com/articles/s41591-019-0503-6 (visited on 03/29/2020) (page 10).

[41] Xavier Didelot, Rory Bowden, Daniel J. Wilson, Tim E. A. Peto, and Derrick W. Crook. "Transforming clinical microbiology with bacterial genome sequencing". en. In: *Nature Reviews Genetics* 13.9 (Sept. 2012). Number: 9 Publisher: Nature Publishing Group, pp. 601–612. ISSN: 1471-0064. DOI: 10.1038/nrg3226. URL: https://www.nature.com/articles/nrg3226 (visited on 11/07/2020) (page 10).

[42] Phelim Bradley, N. Claire Gordon, Timothy M. Walker, Laura Dunn, Simon Heys, Bill Huang, Sarah Earle, Louise J. Pankhurst, Luke Anson, Mariateresa de Cesare, Paolo Piazza, Antonina A. Votintseva, Tanya Golubchik, Daniel J. Wilson, David H. Wyllie, Roland Diel, Stefan Niemann, Silke Feuerriegel, Thomas A. Kohl, Nazir Ismail, Shaheed V. Omar, E. Grace Smith, David Buck, Gil McVean, A. Sarah Walker, Tim E. A. Peto, Derrick W. Crook, and Zamin Iqbal. "Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis". en. In: *Nature Communications* 6.1 (Dec. 2015). Number: 1 Publisher: Nature Publishing Group, p. 10063. ISSN: 2041-1723. DOI: 10.1038/ncomms10063. URL: https://www.nature.com/articles/ncomms10063 (visited on 11/07/2020) (page 10).

[43] Xavier Didelot and Koen B. Pouwels. "Machine-learning-assisted selection of antibiotic prescription". en. In: *Nature Medicine* 25.7 (July 2019). Number: 7 Publisher: Nature Publishing Group, pp. 1033–1034. ISSN: 1546-170X. DOI: 10.1038/s41591-019-0517-0. URL: https://www.nature.com/articles/s41591-019-0517-0 (visited on 11/03/2020) (page 10).

[44] Marcelo Low, Ami Neuberger, Thomas M. Hooton, Manfred S. Green, Raul Raz, Ran D. Balicer, and Ronit Almog. "Association between urinary community-acquired fluoroquinolone-resistant Escherichia coli and neighbourhood antibiotic consumption: a population-based case-control study". English. In: *The Lancet Infectious Diseases* 19.4 (Apr. 2019). Publisher: Elsevier, pp. 419–428. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(18)30676-5. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30676-5/abstract (visited on 11/07/2020) (page 12).

[45] Juan-Carlos Jiménez-Castellanos, Wan Ahmad Kamil Wan Nur Ismah, Yuiko Takebayashi, Jacqueline Findlay, Thamarai Schneiders, Kate J. Heesom, and Matthew B. Avison. "Envelope proteome changes driven by RamA overproduction in Klebsiella pneumoniae that enhance acquired -lactam resistance". en. In: *Journal of Antimicrobial Chemotherapy* 73.1 (Jan. 2018). Publisher: Oxford Academic, pp. 88–94. ISSN: 0305-7453. DOI: 10.1093/jac/dkx345. URL: https://academic.oup.com/jac/article/73/1/88/4364866 (visited on 10/26/2020) (page 12).

[46] Jeroen Berrevoets, James Jordon, Ioana Bica, alexander gimson alexander, and Mihaela van der Schaar. "OrganITE: Optimal transplant donor organ offering using an individual treatment effect". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 20037–20050. URL: https://proceedings.neurips.cc/paper/2020/hash/e7c573c14a09b84f6b7782ce3965f335-Abstract.html (visited on 09/13/2021) (page 14).

[47] Ioana Bica, James Jordon, and Mihaela van der Schaar. "Estimating the Effects of Continuous-valued Interventions using Generative Adversarial Networks". en. In: (Feb. 2020). URL: https://arxiv.org/abs/2002.12326v2 (visited on 09/20/2021) (pages 14, 15).

[48] Ahmed Alaa and Mihaela Schaar. "Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design". en. In: *International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2018, pp. 129–138. URL: https://proceedings.mlr.press/v80/alaa18a.html (visited on 09/20/2021) (pages 14, 32).

[49] Trent Kyono, Ioana Bica, Zhaozhi Qian, and Mihaela van der Schaar. "Selecting Treatment Effects Models for Domain Adaptation Using Causal Knowledge". en. In: (Sept. 2020). URL: https://openreview.net/forum?id=AJY3fGPF1DC (visited on 09/10/2021) (page 15).

[50] Hans-Georg Eichler, Franz Koenig, Peter Arlett, Harald Enzmann, Anthony Humphreys, Frank Pétavy, Brigitte Schwarzer-Daum, Bruno Sepodes, Spiros Vamvakas, and Guido Rasi. "Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation". eng. In: *Clinical Pharmacology and Therapeutics* 107.4 (Apr. 2020), pp. 773–779. ISSN: 1532-6535. DOI: 10.1002/cpt.1638 (page 16).

[51] Brett Beaulieu-Jones, Samuel G. Finlayson, Corey Chivers, Irene Chen, Matthew McDermott, Jaz Kandola, Adrian V. Dalca, Andrew Beam, Madalina Fiterau, and Tristan Naumann. "Trends and Focus of Machine Learning Applications for Health Research". eng. In: *JAMA network open* 2.10 (Oct. 2019), e1914051. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2019.14051 (page 17).

[52] Rodrigo Serafim, José Andrade Gomes, Jorge Salluh, and Pedro Póvoa. "A Comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome Criteria for the Diagnosis of Sepsis and Prediction of Mortality: A Systematic Review and Meta-Analysis". eng. In: *Chest* 153.3 (2018), pp. 646–655. ISSN: 1931-3543. DOI: 10.1016/j.chest.2017.12.015 (page 19).

[53] Robert Hiensch, Jashvant Poeran, Patricia Saunders-Hao, Victoria Adams, Charles A. Powell, Allison Glasser, Madhu Mazumdar, and Gopi Patel. "Impact of an electronic sepsis initiative on antibiotic use and health care facility-onset Clostridium difficile infection rates". eng. In: *American Journal of Infection Control* 45.10 (Oct. 2017), pp. 1091–1100. ISSN: 1527-3296. DOI: 10.1016/j.ajic.2017.04.005 (page 19).

[54] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, and A. Damiano. "The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults". eng. In: *Chest* 100.6 (Dec. 1991), pp. 1619–1636. ISSN: 0012-3692. DOI: 10.1378/chest.100.6.1619 (page 19).

[55] Emanuel P. Rivers, H. Bryant Nguyen, and David Amponsah. "Sepsis: a landscape from the emergency department to the intensive care unit". eng. In: *Critical Care Medicine* 31.3 (Mar. 2003), pp. 968–969. ISSN: 0090-3493. DOI: 10.1097/01.CCM.0000057402.00102.AD (page 19).

[56] Nathan I. Shapiro, Richard E. Wolfe, Richard B. Moore, Eric Smith, Elizabeth Burdick, and David W. Bates. "Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule". eng. In: *Critical Care Medicine* 31.3 (Mar. 2003), pp. 670–675. ISSN: 0090-3493. DOI: 10.1097/01.CCM.0000054867.01688.D1 (page 19).

[57] Michael Oberst, Fredrik D. Johansson, Dennis Wei, Tian Gao, Gabriel Brat, David Sontag, and Kush R. Varshney. "Characterization of Overlap in Observational Studies". In: *arXiv:1907.04138 [cs, stat]* (June 2020). arXiv: 1907.04138. URL: http://arxiv.org/abs/1907.04138 (visited on 10/11/2021) (page 24).

[58] Alexander Franks, Alexander D'Amour, and Avi Feller. "Flexible sensitivity analysis for observational studies without observable implications". In: *arXiv:1809.00399 [stat]* (Jan. 2019). arXiv: 1809.00399. URL: http://arxiv.org/abs/1809.00399 (visited on 10/20/2021) (page 26).

[59] Yixin Wang and David M. Blei. "The Blessings of Multiple Causes". In: *arXiv:1805.06826 [cs, stat]* (Apr. 2019). arXiv: 1805.06826. URL: http://arxiv.org/abs/1805.06826 (visited on 10/20/2021) (page 26).

[60] Peter W G Tennant, Eleanor J Murray, Kellyn F Arnold, Laurie Berrie, Matthew P Fox, Sarah C Gadd, Wendy J Harrison, Claire Keeble, Lynsie R Ranker, Johannes Textor, Georgia D Tomova, Mark S Gilthorpe, and George T H Ellison. "Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations". In: *International Journal of Epidemiology* 50.2 (Apr. 2021), pp. 620–632. ISSN: 0300-5771. DOI: 10.1093/ije/dyaa213. URL: https://doi.org/10.1093/ije/dyaa213 (visited on 09/01/2021) (page 27).

[61] Jason S. Haukoos and Craig D. Newgard. "Advanced Statistics: Missing Data in Clinical Research—Part 1: An Introduction and Conceptual Framework". en. In: *Academic Emergency Medicine* 14.7 (2007). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1553-2712.2007.tb01855.x, pp. 662–668. ISSN: 1553-2712. DOI: 10.1111/j.1553-2712.2007.tb01855.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1553-2712.2007.tb01855.x (visited on 10/23/2020) (page 27).

[62] Rolf H. H. Groenwold. "Informative missingness in electronic health record systems: the curse of knowing". In: *Diagnostic and Prognostic Research* 4.1 (July 2020), p. 8. ISSN: 2397-7523. DOI: 10.1186/s41512-020-00077-0. URL: https://doi.org/10.1186/s41512-020-00077-0 (visited on 11/05/2020) (page 27).

[63] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. "Recurrent Neural Networks for Multivariate Time Series with Missing Values". en. In: *Scientific Reports* 8.1 (Apr. 2018). Number: 1 Publisher: Nature Publishing Group, p. 6085. ISSN: 2045-2322. DOI: 10.1038/s41598-018-24271-9. URL: https://www.nature.com/articles/s41598-018-24271-9 (visited on 11/05/2020) (page 27).

[64] Jinsung Yoon, William R. Zame, and Mihaela van der Schaar. "Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks". In: *IEEE Transactions on Biomedical Engineering* 66.5 (May 2019). Conference Name: IEEE Transactions on Biomedical Engineering, pp. 1477–1490. ISSN: 1558-2531. DOI: 10.1109/TBME.2018.2874712 (page 28).

[65] Brett K. Beaulieu-Jones, Samuel G. Finlayson, William Yuan, Russ B. Altman, Isaac S. Kohane, Vinay Prasad, and Kun-Hsing Yu. "Examining the Use of Real-World Evidence in the Regulatory Process". eng. In: *Clinical Pharmacology and Therapeutics* 107.4 (Apr. 2020), pp. 843–852. ISSN: 1532-6535. DOI: 10.1002/cpt.1658 (page 32).

[66] Conceição Granja, Wouter Janssen, and Monika Alise Johansen. "Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature". eng. In: *Journal of Medical Internet Research* 20.5 (2018), e10235. ISSN: 1438-8871. DOI: 10.2196/10235 (page 33).

[67] Lucas M. Fleuren, Thomas L. T. Klausch, Charlotte L. Zwager, Linda J. Schoonmade, Tingjie Guo, Luca F. Roggeveen, Eleonora L. Swart, Armand R. J. Girbes, Patrick Thoral, Ari Ercole, Mark Hoogendoorn, and Paul W. G. Elbers. "Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy". eng. In: *Intensive Care Medicine* 46.3 (2020), pp. 383–400. ISSN: 1432-1238. DOI: 10.1007/s00134-019-05872-y (page 33).

[68] Robert M Wachter. "Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England". en. In: *Making IT Work* (), p. 71 (page 34).

[69] Hugh Alderwick and Jennifer Dixon. "The NHS long term plan". en. In: *BMJ* 364 (Jan. 2019). Publisher: British Medical Journal Publishing Group Section: Editorial. ISSN: 0959-8138, 1756-1833. DOI: 10.1136/bmj.l84. URL: https://www.bmj.com/content/364/bmj.l84 (visited on 11/04/2020) (page 34).

7 LABMARCS

LABMARCS: **LAB**oratory **MA**rkers of **C**OVID-19 **S**everity - Bristol Cohort

Mr Edward Barker[a,*], Dr Ranjeet Bhamber[a], Dr Louis McGregor[a], Dr Mahableshwar Albur[b], Dr Philip Williams[c], Dr Matt Thomas[b], Dr Stefan Gurney[c], Dr Catherine Hyams[a,b], Prof Andrew Dowsey[a], Prof Jonathan Sterne[a], Dr Katy Turner[a], Dr Chris McWilliams[a]

[a] *University Of Bristol*
[b] *Southmead Hospital, North Bristol NHS Trust*
[c] *University Hospitals Bristol and Weston NHS Trust UK*

Abstract

Objectives: (1) To explore independent associations of routinely measured physiological and blood parameters at hospital admission with disease severity (ICU admission or death at 28 days from hospital admission), adjusting for demographics and comorbidities. (2) To develop a prediction model for severe COVID-19 outcomes at 28 days in order to generate a simple severity scoring system so that the frontline clinicians can prioritise resources to the most critically ill patients to improve their outcome

Methods: A retrospective non-interventional (observational) cohort study of patients with COVID-19, using routine laboratory tests performed as part of the patients' standard of care.

Results: Testing

Conclusions: Testing

Keywords:

Background

As of 30 April 2021, there have been approximately 150 million confirmed cases of COVID-19 disease worldwide, with 4.4 million cases in the UK, resulting in over 127,000 deaths (WHO Coronavirus (COVID-19) Dashboard — WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, n.d.). COVID-19 has a wide spectrum of clinical features ranging from asymptomatic to severe systemic illness with a significant attributable mortality, while clinical manifestations are variable especially in the most vulnerable groups and immunocompromised people. The majority of people (85%) with COVID-19 will have mild or no symptoms. A proportion (10%) of them suffer from a severe infection needing hospital admission for supportive care such as oxygen and a minority (5%) will require intensive care therapy such as mechanical ventilation. Clinical manifestations are variable, especially in the most vulnerable groups and immunocompromised people [1]

Of those that require hospitalisation, early identification of those likely to deteriorate; requiring transfer to an intensive care unit (ICU) or who may die, is vital for clinical decision making. Healthcare systems across the world including in highly developed countries such as in Europe and the USA have been faced with immense challenges in terms of capacity and resources to manage this pandemic, which may continue as we enter the relaxation of lock down measures including opening of schools and businesses. Numerous studies have identified clinical factors associated with requirement for intensive care and death among those with COVID-19 infection. Published prediction models to date have evaluated case-level factors that might predict poor outcomes (critical illness or death). A recent living systematic review looked at diagnostic and prognostic models [2]. It identified 107 prognostic models for predicting mortality risk; the same type shown here in this paper, the majority of which looked at vital signs, age, comorbidities and image features such as chest X-rays. Models which included a broad range of variables concerning coinfection, biochemical factors outside of C-reactive protein, and other haematological factors on an individual patient level were not common. All of these prognostic models showed a high or unclear risk of bias, did not describe the target population or care setting, and only 11 were externally validated by a calibration plot [2]. It was also uncommon for models to include efforts to convert prediction models into parsimonious, simple, score-based tools that can be used easily for risk stratification in clinical settings. In the context of hospital based cases of COVID-19, such rules might have important implications for risk stratification of patients, streamlining decisions around transfer to ICU or discharge. This study therefore aims to analyse all the available laboratory values of different metabolic pathways affected by COVID-19 infection over a comprehensive range of patient variables. We aim to understand their inter-relatedness and impact on key clinical outcomes such as need for admission to hospital, transfer of care to ITU and death. The two main objectives of this study are: (1) To explore independent associations of routinely measured physiological and blood parameters at hospital admission with disease severity (ICU admission or death at 28 days from hospital admission), adjusting for demographics and comorbidities.

*Corresponding author.
E-mail address: e.barker@bristol.ac.uk

(2) To develop a prediction model for severe COVID-19 outcomes at 28 days in order to generate a simple severity scoring system so that the frontline clinicians can prioritise resources to the most critically ill patients to improve their outcome. Also help them to discharge patients home to free the hospital bed capacity for the most severely ill patients.

Derangements of homeostasis in COVID-19

COVID-19 has proven to be a challenging disease with multisystem affection resulting in the derangements of homeostasis in addition to systemic inflammatory response syndrome, mimicking sepsis. The derangements in homeostasis include pulmonary, cardiovascular, coagulation, haematological, oxygenation, renal and fluid balance. They are reflected in variety of laboratory test abnormalities measured as a part of routine clinical care in the hospital setting on a regular basis. Initial case series from Wuhan have shown that few of these laboratory abnormalities such as D-dimer, neutrophil-lymphocyte ratio, ferritin are associated with poor clinical outcomes [3]. However, they reflect a pro-inflammatory response to any infection and are commonly seen in bacterial sepsis, and hence not specific to COVID

Derangements of coagulation cascade in COVID-19

A significant proportion of patients with Covid-19 have a profound hypercoagulable state. Arterial and venous thrombotic events are not uncommon. Abnormalities in coagulation screening tests, including a prolonged activated partial-thromboplastin time (aPTT), have been reported in patients with Covid-19. This has raised some concerns as well as confusion amongst clinicians regarding anticoagulation. A recent study from a teaching hospital in London (UK) included 216 patients with COVID-19 who had coagulation screening, 44 (20%) were found to have a prolonged aPTT [4]. Most patients with Covid-19 in that cohort who were admitted to the hospital with a prolonged aPTT were positive for lupus anticoagulant (91%) and often had an associated factor XII deficiency. When compared with a historic cohort of 540 patients, the percentage of specimens that were positive for lupus anticoagulant was significantly higher among the patients with Covid-19 (P¡0.001). However the authors concluded that the role, if any, of lupus anticoagulant in the pathogenesis of COVID-19 thrombosis is unknown and suggested further research in this area [4]. Retrospective studies from China showed poor clinical outcomes in patients with COVID19 presenting with abnormal coagulation tests [5, 6]. Another cohort study from China looked at prediction of severity of illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. There were some clear signals of such an association in that study [7].

Inflammatory markers in COVID-19

It has been shown that inflammation plays a key role in pathogenesis of COVID-19. In these patients release of pro-inflammatory cytokines such as IL-6 results in activation of inflammatory process producing lung injury [8]. A metaanalysis involving 12 studies from China with a total cohort of ¿700 patients showed that elevation in inflammatory markers such as C-reactive protein, IL-6 and Procalcitonin are associated with poor clinical outcome [9]. A prospective single centre study from Wuhan (China) involving 150 COVID-19 patients looked at predictors of a fatal outcome [10]. Apart from old age, presence of underlying comorbidities and elevated inflammatory indicators in the blood were associated with poor outcome. The authors hypothesised that COVID-19 mortality might be due to virus-activated "cytokine storm syndrome" or fulminant myocarditis. A case series from Singapore also echoed similar association of an elevated CRP with hypoxemic respiratory failure [11].

Blood counts in COVID-19

Systemic inflammatory response can lead to abnormalities in circulating blood cells in the form of neutrophilia with lymphopenia, as a response of the innate immune system to COVID-19. The neutrophil-to-lymphocyte ratio (NLR), the proportion of absolute neutrophil count to lymphocytes has been suggested as a marker of COVID severity. An elevated NLR (¿3.3) is shown to be associated with poor prognosis in COVID-19 [12].

Oxygenation/ventilator derangements in COVID-19

Mechanism of severe acute respiratory syndrome caused by coronavirus-induced acute lung injury have been reviewed extensively against SARS CoV1 and MERSA CoV [13]. Hypoxaemia is a common finding in severe COVID-19 patients and is the primary reason for admission to intensive care unit. Initial studies from Wuhan (China) showed mixed pathophysiological changes in the lungs as a cause of ARDS resulting in type 1 respiratory failure [14]. Subsequent case studies across the world have shown varying picture in terms of prognostic markers of mechanically ventilated patients with COVID-19 [15, 16, 17].

Patients and methods

Study cohort

The training cohort (n = 1171) was defined as all adult inpatients testing positive for severe acute respiratory syndrome coronavirus 2 (SARS-Cov2) by reverse transcription polymerase chain reaction (RT-PCR) between 9th March 2020 and 17th October 2020 at North Bristol NHS trust (NBT) and University Hospitals Bristol NHS Foundation Trust (UHBW). For external validation purposes, we used X cohorts

Demographics
Data source

Pseudonymised Severn Pathology data from WinPath linking laboratory markers to key clinical outcomes. Severn Pathology carries out testing for three trusts including UHBW, NBT and Bath. A system wide data search was conducted on the LIMS for all patients who have tested positive for COVID-19 infection. All pathology data was collected including bacteriology, virology, mycology, haematology and biochemistry collected as part of clinical care of COVID patients. The serial data set allowed us to identify transfer to ICU and mortality including the timelines.

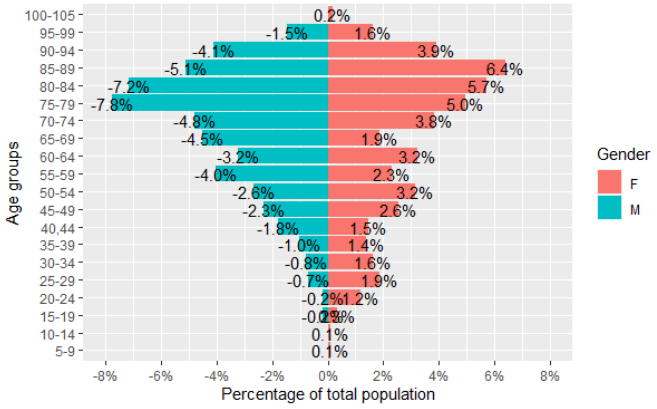


Figure 1: Distribution of hospitalized patients with coronavirus disease 2019 (COVID-19) stratified by age and gender

Setting and inclusion criteria

All adult patients presenting/admitted to hospitals across Bristol (NBT and UHBW) and tested positive for SARS-CoV-2 by PCR at Severn Pathology Laboratory Information Management System (LIMS).

Exclusion criteria

- Paediatric patients (age <18 years of age)
- Staff / healthcare worker house-hold contacts

Covariates

Predictors included host factors, clinical severity indices, microbiological factors, immunological factors, haematological factors and biochemical factors. Full list of data items included in the Supplementary Material.

Microbiological Factors Immunological factors Haematological factors Biochemical factors Demographics and comorbidities

Outcomes

For all sites, the primary outcome was severe COVID-19 disease at 30 days following hospital admission, categorised as transfer to the ICU/death (WHO-COVID-19 Outcomes Scales 6–8) vs. not transferred to the ICU/death (scales 3–5). For nosocomial patients (patients with symptom onset after hospital admission), the endpoint was defined as 28 days after a positive COVID-19 test result. Dates of hospital admission, symptom onset, ICU transfer, and death were extracted from electronic health records or ascertained manually by a clinician.

The secondary outcome is a COVID-19 specific disease severity score.

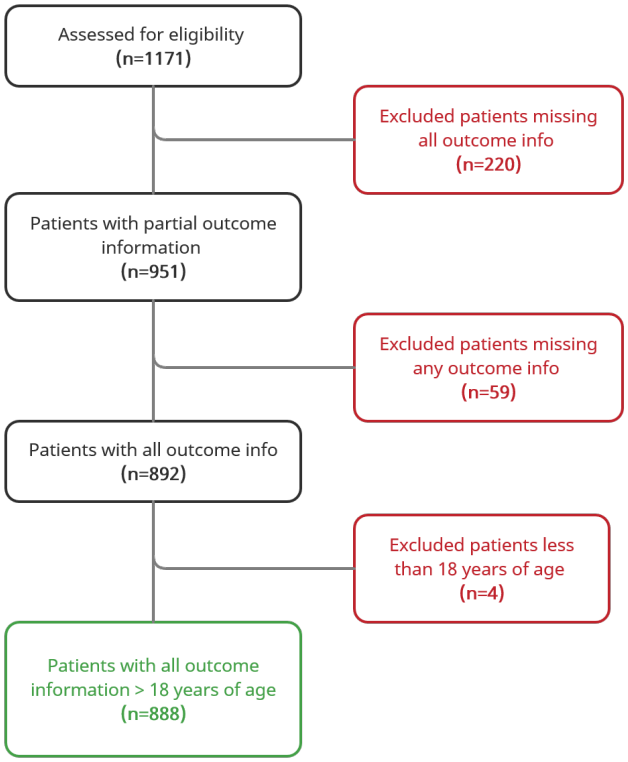


Figure 2: Flowchart of exclusion and inclusion criteria

Data extraction

The key clinical outcome data was collected from WinPath and Sunquest-ICE laboratory information management systems. Severn Infection Sciences laboratory Polymerase Chain Reaction (PCR) tests for COVID-19 were extracted. All COVID-19 negative patients were excluded. All laboratory markers including clinical outcomes from Laboratory Information Management Systems (LIMS) were extracted and the final dataset was anonymized with no patient identifying data to link back.

Data processing

Outcomes came as separate comma separated value (CSV) files, with unique identifiers for each patient, and one row per admission. Changed date format Outcome for NBT listed as ”Patient died” if the patient died. Created a new variable for date of death set as the date of discharge when the outcome was ”Patient died” Created a new variable ”Site” to specify the trust where the data originated. UHB only had admission date and length of stay, and new discharge date variable was created based on the length of stay added to the admission date. Also created a new variable ”Site” to specify the trust where the data originated. Merged the outcome data from UHB, NBT and Weston with the now common variables Merged the demographics data frame (Age and Gender) with the outcomes data frame by ID to create a ”main” data frame. Removed those patients with missing admission date. Removed those patients with missing discharge date. Removed those patients who met the exclusion criteria (Age ¡ 18) Created a new binary variable for whether

or not a patient had died based on the presence of a death date, and added this to the main data frame. Created a new binary variable for whether or not a patient went to ICU based on the presence of an ICU admission date, and added this to the main data frame. Finally, created a new severe outcome variable for whether or not a patient either died or went to the ICU, and added this to the main data frame.

Covid PCR tests came as a separate CSV file, and included all COVID-19 positive tests for all patients who had ever tested positive for COVID-19 between the dates specified earlier. Our test of interest was the first COVID-19 positive test. For each unique ID within this data frame, we kept the first positive COVID-19 test. Finally, we merged this first positive test date with the main data frame.

There were multiple admissions associated with a number of the patients in the cohort. All routine laboratory test information and outcomes were gathered based on the dates of the first COVID-19 positive test, and so all data including and from the date of the first COVID-19 positive test was included. As such, we only kept admission dates relevant to the first COVID-19 positive test, and excluded admission dates for subsequent admissions.

Each laboratory variable came as a separate CSV file including the IDs of the patients in the cohort.

Maintaining variables as continuous would lead to the greatest amount of information and statistical power from the data. However, many of the continuous variables of interest in this study would be considered "worse" in a clinical sense whether they are high or low. There is an opportunity to dichotomise these continuous variables, however this comes at a cost. Much information is lost, the statistical power to detect a relation is reduced, it increases the risk of a positive result being a false positive, and it may underestimate the extent of variation in outcome between groups [18]. Considering our sample size is relatively small, the addition of new variables would limit how many laboratory variables we could take into account in our final model.

Many variables had reference ranges for clinical use, though these ranges were for a normal population, and not specific to COVID-19 positive patients.

Missing data

Statistical analyses

All continuous parameters were winsorized (at 1% and 99%) and scaled (mean = 0; standard deviation = 1) to facilitate interpretability and comparability [**?**]. Logarithmic or square root transformations were applied to skewed parameters so that the parameter would have a distribution that was approximately normal. Variables related to test results were categorised by clinical significance as either 'Normal' or 'Abnormal'. To explore independent associations of blood and physiological parameters with 28-day ICU/death, we used logistic regression with Firth's bias reduction method [19]. Each parameter was tested independently, adjusted for age and sex (model 1), and then additionally adjusted for comorbidities (model 2). P values were adjusted using the Benjamini-Hochberg procedure to keep the false discovery rate (FDR) at 5% [20].

For our clinical prediction rule we utilised regularised logistic regression with a least absolute shrinkage and selection operator (LASSO) estimator that shrinks parameters according to their variance, reduces overfitting, and enables automatic variable selection [21]. The optimal degree of regularisation was determined by identifying a tuning parameter λ using cross-validation. To avoid overfitting and to reduce the number of false-positive predictors, λ was selected to give a model with an area under the receiver operating characteristic curve (AUC) one standard error below the 'best' model. To evaluate the predictive performance of our model on new cases of the same underlying population (internal validation), we performed repeated nested cross-validation (10-folds the for inner loop; 10-folds and 1000 repeats for the outer loop). Discrimination was assessed using AUC and Brier score. Missing feature information was imputed using k-nearest neighbour (kNN) imputation (k = 5) for variables deemed missing at random . All variables which were not deemed to be missing at random were catagorical in nature and so we added 'No Value Available' as a separate category All steps (feature selection, winsorizing, scaling, and kNN imputation) were incorporated within the model development and selection process to avoid data leakage that would otherwise result in optimistic performance measures [22]. All analyses were conducted with R version 4.0.1 using the glmnet [37] and coxnet [REF] packages. We evaluated the transportability of the derived regularised logistic regression model in external validation samples from X (n = 888) ... Validation used LASSO logistic regression models trained on the original combined training sample. Models were assessed in terms of discrimination (AUC, sensitivity, specificity, Brier score), calibration, and clinical utility (decision curve analysis, number needed to evaluate) [32, 39]. Moderate calibration was assessed by plotting model-predicted probabilities (x-axis) against observed proportions (y-axis) with locally estimated scatterplot smoothing (LOESS) and logistic curves [40]. Clinical utility was assessed using decision curve analysis where 'net benefit' was plotted against a range of threshold probabilities. Unlike diagnostic performance measures, decision curves incorporate preferences of the clinician and patient. The threshold probability (pt) is where the expected benefit of treatment is equal to the expected benefit of avoiding treatment [41]. Net benefit was calculated by counting the number of true positives (predicted risk ¿ pt and experienced severe COVID-19 outcome) and false positives (predicted risk ¿ pt but did not experience severe COVID-19 outcome) and using the below formula: Net benefit = True positivesN-False positivesN*pt1-ptNet benefit=True positivesN-False positivesN*pt1-pt Our model was developed as a screening tool, to identify at hospital admission patients at risk of more severe outcomes. The intended treatment for patients with a positive result from this model would be further examination by a clinician, who would make recommendations regarding appropriate treatment (e.g. earlier transfer to the ICU, intensive monitoring, treatment). We compared the decision curve from our model to two extreme cases of 'treat none' and 'treat all'. The 'treat none' (i.e. routine management) strategy implies that no patients would be selected for further examination by a clinician; the 'treat all' strategy (i.e. intensive

management) implies that all patients would undergo further assessment. A model is clinically beneficial if the model-implied net benefit is greater than either the 'treat none' or 'treat all' strategies. Since the intended strategy involves a further examination by a clinician, and is therefore low risk, our emphasis throughout is on avoiding false negatives (i.e. failing to detect a severe case) at the expense of false positives. We therefore used a threshold of 30% to calculate sensitivity and specificity. This gave a better balance of sensitivity vs. specificity and reflected the clinical preference to avoid false negatives for the proposed screening tool.

Sensitivity analyses

We conducted five sensitivity analyses. First, to explore the ability to predict shorter-term severe COVID-19 outcome, we developed models for ICU transfer/death at 3 and 7 days following hospital admission. All steps described above were repeated, including training (feature selection) and external validation. Second, different imputation assumptions were made, with all values imputed as if missing at random and X methodology used instead. Third, we adjusted the threshold at which we would drop a variable due to missing data from 30% to 20% and 40% respectively. Fourth, we examined using data from only the first day of admission and data from up to 5 days following admission. Finally, we considered the differences between community-acquired vs. nosocomial infection, we repeated all models after excluding the X nosocomial patients or the community-acquired infections respectively.

Results

Model development

Model specification

Model performance

Stability Analysis

Validation

External test data.

Split test data.

Discussion

Summary of results

Strengths and limitations

Limitations: Small cohort in one area of England where case numbers have varied widely. Results in less precise parameter estimates for prediction models (less power/smaller sample size) and reduced generalisability of the model to other settings. Strengths: Granularity of laboratory data available linked to clinical outcomes.

Comparison with other research

Implications for policy/practice

Prediction model giving the absolute probability for the risk of mortality will assist frontline clinicians in their assessments of COVID-19 patients in terms of severity of illness and potential outcomes. This will help clinical decision making for hospital admission or discharge or transfer to an intensive care unit. Identification of key laboratory parameters of severity of COVID-19 in predicting the outcome and prioritise resources to the most critically ill patients to improve their outcome.

Future research

Validation of the prediction model.

Conclusions

Ethics approval

The study underwent a rigorous ethical and regulatory approval process and a favourable opinion was gained from an REC (UK Health Departments Research Ethics Service NHS REC).

Funding

Authors contributions

Declaration of competing interest

Acknowledgements

Supplementary materials

References

References

[1] A. B. Docherty, E. M. Harrison, C. A. Green, H. Hardwick, R. Pius, L. Norman, K. A. Holden, J. M. Read, F. Dondelinger, G. Carson, L. Merson, J. Lee, D. Plotkin, L. Sigfrid, S. Halpin, C. Jackson, C. Gamble, P. W. Horby, J. S. Nguyen-Van-Tam, I. Investigators, J. Dunning, P. J. M. Openshaw, J. K. Baillie, M. G. Semple, Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol, medRxiv (2020) 2020.04.23.20076042doi:10.1101/2020.04.23.20076042.

[2] L. Wynants, B. V. Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. J. Bonten, D. L. Dahly, J. A. Damen, T. P. A. Debray, V. M. T. de Jong, M. D. Vos, P. Dhiman, M. C. Haller, M. O. Harhay, L. Henckaerts, P. Heus, M. Kammer, N. Kreuzberger, A. Lohmann, K. Luijken, J. Ma, G. P. Martin, D. J. McLernon, C. L. A. Navarro, J. B. Reitsma, J. C. Sergeant, C. Shi, N. Skoetz, L. J. M. Smits, K. I. E. Snell, M. Sperrin, R. Spijker, E. W. Steyerberg, T. Takada, I. Tzoulaki, S. M. J. van Kuijk, B. C. T. van Bussel, I. C. C. van der Horst, F. S. van Royen, J. Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K. G. M. Moons, M. van Smeden, Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal, BMJ 369 (2020) m1328. doi:10.1136/bmj.m1328.

[3] C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Y. Song, Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China, JAMA internal medicine 180 (7) (2020) 934–943. doi:10.1001/jamainternmed.2020.0994.

[4] L. Bowles, S. Platton, N. Yartey, M. Dave, K. Lee, D. P. Hart, V. MacDonald, L. Green, S. Sivapalaratnam, K. J. Pasi, P. MacCallum, Lupus Anticoagulant and Abnormal Coagulation Tests in Patients with Covid-19, New England Journal of Medicine 383 (3) (2020) 288–290. doi:10.1056/NEJMc2013656.

[5] N. Tang, D. Li, X. Wang, Z. Sun, Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia, Journal of thrombosis and haemostasis: JTH 18 (4) (2020) 844–847. doi:10.1111/jth.14768.

[6] H. Han, L. Yang, R. Liu, F. Liu, K.-L. Wu, J. Li, X.-H. Liu, C.-L. Zhu, Prominent changes in blood coagulation of patients with SARS-CoV-2 infection, Clinical Chemistry and Laboratory Medicine 58 (7) (2020) 1116–1120. doi:10.1515/cclm-2020-0188.

[7] X. Bi, Z. Su, H. Yan, J. Du, J. Wang, L. Chen, M. Peng, S. Chen, B. Shen, J. Li, Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count, Platelets 31 (5) (2020) 674–679. doi:10.1080/09537104.2020.1760230.

[8] F. Liu, L. Li, M. Xu, J. Wu, D. Luo, Y. Zhu, B. Li, X. Song, X. Zhou, Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19, Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology 127 (2020) 104370. doi:10.1016/j.jcv.2020.104370.

[9] G. Vaseghi, M. Mansourian, R. Karimi, K. Heshmat-Ghahdarijani, P. Rouhi, M. Shariati, S. H. Javanmard, Inflammatory markers in Covid-19 Patients: A systematic review and meta-analysis, medRxiv (2020) 2020.04.29.20084863doi:10.1101/2020.04.29.20084863.

[10] Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, Intensive Care Medicine 46 (5) (2020) 846–848. doi:10.1007/s00134-020-05991-x.

[11] B. E. Young, S. W. X. Ong, S. Kalimuddin, J. G. Low, S. Y. Tan, J. Loh, O.-T. Ng, K. Marimuthu, L. W. Ang, T. M. Mak, S. K. Lau, D. E. Anderson, K. S. Chan, T. Y. Tan, T. Y. Ng, L. Cui, Z. Said, L. Kurupatham, M. I.-C. Chen, M. Chan, S. Vasoo, L.-F. Wang, B. H. Tan, R. T. P. Lin, V. J. M. Lee, Y.-S. Leo, D. C. Lye, Singapore 2019 Novel Coronavirus Outbreak Research Team, Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore, JAMA 323 (15) (2020) 1488–1494. doi:10.1001/jama.2020.3204.

[12] J. Liu, Y. Liu, P. Xiang, L. Pu, H. Xiong, C. Li, M. Zhang, J. Tan, Y. Xu, R. Song, M. Song, L. Wang, W. Zhang, B. Han, L. Yang, X. Wang, G. Zhou, T. Zhang, B. Li, Y. Wang, Z. Chen, X. Wang, Neutrophil-to-Lymphocyte Ratio Predicts Severe Illness Patients with 2019 Novel Coronavirus in the Early Stage, medRxiv (2020) 2020.02.10.20021584doi:10.1101/2020.02.10.20021584.

[13] L. E. Gralinski, A. Bankhead, S. Jeng, V. D. Menachery, S. Proll, S. E. Belisle, M. Matzke, B.-J. M. Webb-Robertson, M. L. Luna, A. K. Shukla, M. T. Ferris, M. Bolles, J. Chang, L. Aicher, K. M. Waters, R. D. Smith, T. O. Metz, G. L. Law, M. G. Katze, S. McWeeney, R. S. Baric, Mechanisms of severe acute respiratory syndrome coronavirus-induced acute lung injury, mBio 4 (4). doi:10.1128/mBio.00271-13.

[14] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao, F.-S. Wang, Pathological findings of COVID-19 associated with acute respiratory distress syndrome, The Lancet Respiratory Medicine 8 (4) (2020) 420–422. doi:10.1016/S2213-2600(20)30076-X.

[15] M. Arentz, E. Yim, L. Klaff, S. Lokhandwala, F. X. Riedo, M. Chong, M. Lee, Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State, JAMA 323 (16) (2020) 1612–1614. doi:10.1001/jama.2020.4326.

[16] N. Wilson, A. Kvalsvig, L. T. Barnard, M. G. Baker, Case-Fatality Risk Estimates for COVID-19 Calculated by Using a Lag Time for Fatality - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases journal - CDCdoi:10.3201/eid2606.200320.

[17] H. Barrasa, J. Rello, S. Tejada, A. Martín, G. Balziskueta, C. Vinuesa, B. Fernández-Miret, A. Villagra, A. Vallejo, A. San Sebastián, S. Cabañes, S. Iribarren, F. Fonseca, J. Maynar, Alava COVID-19 Study Investigators, SARS-CoV-2 in Spanish Intensive Care Units: Early experience with 15-day survival in Vitoria, Anaesthesia, Critical Care & Pain Medicine 39 (5) (2020) 553–561. doi:10.1016/j.accpm.2020.04.001.

[18] D. G. Altman, P. Royston, The cost of dichotomising continuous variables, BMJ : British Medical Journal 332 (7549) (2006) 1080.

[19] G. Heinze, M. Ploner, D. Dunkler, H. Southworth, Firth's bias reduced logistic regression, R package version 1.

[20] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological) 57 (1) (1995) 289–300.

[21] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

[22] M. Kuhn, K. Johnson, Data Pre-processing, in: M. Kuhn, K. Johnson (Eds.), Applied Predictive Modeling, Springer, New York, NY, 2013, pp. 27–59. doi:10.1007/978-1-4614-6849-3_3.