

**MACHINE LEARNING APPROACHES TO OPTIMISE THE
MANAGEMENT OF PATIENTS WITH SEPSIS**

PHD THESIS

by:

Dr Matthieu Komorowski

Department of Surgery and Cancer and Department of Bioengineering

Imperial College London

London SW7 2AZ

United Kingdom

Corrected report submitted on 06/02/2019 as part of the requirements for the award of

PhD in Medicine, Imperial College London, 2015-2019

Declaration of Originality

I hereby declare that this PhD thesis is my own work, and it does not contain other people's work without this being stated.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Glossary

AIC	Akaike Information Criterion
AUROC	Area Under the Receiver Operator Curve
BIC	Bayesian Information Criterion
BRC	Biomedical Research Centres
BUN	Blood Urea Nitrogen
CCHIC	Critical Care Health Informatics Collaborative
EGDT	Early-Goal Directed Therapy
EHR	Electronic Health Record
eRI	e-Research Institute
FiO ₂	Fraction of oxygen - inspired
GDPR	General Data Protection Regulation
HCOPE	High-Confidence Off-policy Evaluation
ICD-9	International Classification of Diseases version 9;
ICNARC	Intensive Care National Audit & Research Centre
ICU	Intensive Care Unit
IQR	Inter Quartile Range
IS	Importance Sampling
LB	Lower Bound
MAP	Mean Arterial Pressure
MDP	Markov Decision Process
MIMIC-III	Medical Information Mart for Intensive Care, version 3
NHIC	NIHR Health Informatics Collaborative
NIHR	National Institute of Health Research
OASIS	Oxford Acute Severity of Illness Score
OPE	Off-policy Policy Evaluation
PCA	Principal Component Analysis
PDIS	Per-Decision Importance Sampling
PDWIS	Per-Decision Weighted Importance Sampling
POMDP	Partially Observable Markov Decision Process
RCT	Randomized Controlled Trial
RL	Reinforcement Learning
qSOFA	quick Sequential Organ Failure Assessment
SD	Standard Deviation
SEM	Standard Error of the Mean
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment
SQL	Structured Query Language
SSC	Surviving Sepsis Campaign
TD learning	Temporal Difference learning
t-SNE	t-Distributed Stochastic Neighbor Embedding
UB	Upper bound
WIS	Weighted Importance Sampling

Summary of Notation

An MDP is defined as a tuple $\{S, A, T, r, \gamma\}$ with:

S a set of states

$A_t \sim \pi(\cdot | S_t)$ a set of actions

$T : S \times S \times A \rightarrow [0,1]$ a probability mass function defining $S_{t+1} \sim T(\cdot | S_t, A_t)$

$r : S \times A \rightarrow [-100,100]$ a bounded reward function

$\gamma \in [0,1]$ a discount factor

H is a trajectory of length L , a state-action history

D a set of n historical trajectories

$|D|$ the number of trajectories in D

k the number of states or clusters in the model

C_i the cluster number i

$\pi(a, s)$ any policy defining $p(a | \pi, s)$

π_b a behaviour policy, a clinicians' policy

π_e an evaluation policy, an AI policy

π^* an optimal policy

$V(\pi)$ a state-value function under π

$Q(s, a)$ an action-value function

α a learning rate

Abstract

The goal of this PhD was to generate novel tools to improve the management of patients with sepsis, by applying machine learning techniques on routinely collected electronic health records. Machine learning is an application of artificial intelligence (AI), where a machine analyses data and becomes able to execute complex tasks without being explicitly programmed. Sepsis is the third leading cause of death worldwide and the main cause of mortality in hospitals, but the best treatment strategy remains uncertain. In particular, evidence suggests that current practices in the administration of intravenous fluids and vasopressors are suboptimal and likely induce harm in a proportion of patients. This represents a key clinical challenge and a top research priority.

The main contribution of the research has been the development of a reinforcement learning framework and algorithms, in order to tackle this sequential decision-making problem. The model was built and then validated on three large non-overlapping intensive care databases, containing data collected from adult patients in the U.S.A and the U.K. Our agent extracted implicit knowledge from an amount of patient data that exceeds many-fold the life-time experience of human clinicians and learned optimal treatment by having analysed myriads of (mostly sub-optimal) treatment decisions. We used state-of-the-art evaluation techniques (called high confidence off-policy evaluation) and demonstrated that the value of the treatment strategy of the AI agent was on average reliably higher than the human clinicians. In two large validation cohorts independent from the training data, mortality was the lowest in patients where clinicians' actual doses matched the AI policy. We also gained insight into the model representations and confirmed that the AI agent relied on clinically and biologically meaningful parameters when making its suggestions. We conducted extensive testing and exploration of the behaviour of the AI agent down to the level of individual patient trajectories, identified potential sources of inappropriate behaviour and offered suggestions for future model refinements.

If validated, our model could provide individualized and clinically interpretable treatment decisions for sepsis that may improve patient outcomes.

Table of Contents

Declaration of Originality	2
Copyright Declaration.....	2
Glossary	3
Summary of Notation.....	4
Abstract.....	5
Chapter 1: Background	9
Definitions.....	9
Sepsis epidemiology	10
Sepsis pathophysiology.....	12
Current state-of-the-art sepsis management.....	14
The opportunity for machine learning in decision-making in healthcare	20
Introduction to reinforcement learning and notation	21
Reinforcement learning algorithms.....	24
Clinicians' policy evaluation	24
Optimal policy estimation.....	25
Challenges of model evaluation.....	27
Why is it so difficult?.....	27
Formalising off-policy evaluation.....	29
Related work.....	30
Hypotheses and objectives.....	31
Overview of the approach.....	33
Chapter 2: Data extraction and pre-processing.....	34
The datasets.....	34
Data requirements	36
Defining the patient cohort	37
Inclusion criteria	37
Exclusion criteria	37
Implementation	38
Exclusion of eRI hospitals with low data availability.....	39
Patient inclusion flow diagrams	41
Data extraction	43
Data preparation.....	44
List of model features	46
Description of the datasets	48
Discussion on dataset preparation and patient cohort definition	51

Chapter summary	52
Chapter 3: Model construction.....	53
Choice of model.....	53
Defining the elements of the Markov decision process	55
Defining states	55
Defining actions	64
Transition matrix & Markov property	66
Defining reward and penalty.....	71
Patient trajectories.....	72
Discount factor gamma	73
RL algorithms	74
Discussion on model construction	75
Chapter summary	76
Chapter 4: Model evaluation and policy selection.....	77
Determining optimal actions for test records	77
Model-based off-policy evaluation	78
Importance sampling-based off-policy evaluation.....	81
Selecting the best model using high confidence off-policy evaluation (HCOPE).....	85
Other policy evaluation methods	86
Visual methods.....	86
Direct policy assessment methods	87
Model interpretability: the “right to explanation”.....	92
Further results	95
Variability of AI policy in different models	95
Qualitative comparison of the clinicians’ and AI policies: “internal validation”	96
Behaviour policy estimation	97
Capture of clinical concepts and diagnoses within the states.....	99
Mortality prediction	100
Discussion on model evaluation	103
Chapter summary	104
Chapter 5: Model testing on eRI and CCHIC	105
Model testing on eRI.....	105
Value of the clinicians’ and AI policies.....	105
Qualitative comparison of the clinicians’ and AI policies	108
Relationship between the doses received and mortality	110
Exploration of individual trajectories	111
Interpretability: relative feature importance.....	126

Model testing on CCHIC	128
Chapter summary	131
Chapter 6: Discussion	132
Summary of findings.....	132
Strengths of the approach.....	134
Limitations of the approach	135
Future directions for model development	137
Towards clinical use	140
Conclusion	142
Acknowledgements.....	143
References.....	144
Appendices.....	156
Code availability	156
Research output as of February 2019.....	156
Peer-reviewed papers	156
Book and book chapters.....	157
Abstracts	157
Reviewer activity	158
Prizes and awards.....	159
Invited talks.....	159
Oral Presentations for submitted abstracts.....	160
Permissions to reuse.....	162

Chapter 1: Background

Definitions

Sepsis is defined as a severe infection leading to life-threatening acute organ dysfunction (Singer M, Deutschman CS, Seymour C, & et al, 2016). Sepsis differs from a “simple” infection in the sense that it represents an aberrant or dysregulated host response to the infection and that it requires the presence of organ dysfunction. In the new “sepsis-3” definition, organ dysfunction is defined, in the ICU, by an increase in the Sequential Organ Failure Assessment (SOFA) score from baseline of 2 points or more (Figure 1) (J. L. Vincent et al., 1996). The sepsis-3 definition replaced in 2016 the 15-year old sepsis-2 criteria (Levy et al., 2003). In using the sepsis-3, we adhered to the current international definition, which was confirmed to be superior to Systemic Inflammatory Response Syndrome (SIRS) or quick Sequential Organ Failure Assessment (qSOFA) for sepsis identification in ICU patients (Raith et al., 2017).

Septic shock is a subset of sepsis in which cellular and circulatory abnormalities are severe enough to significantly increase the risk of mortality. It is defined by the presence of persistent hypotension despite adequate fluid volume resuscitation requiring vasopressors to maintain a mean arterial pressure greater than or equal to 65 mmHg and blood lactate greater than or equal to 2 mmol/L.

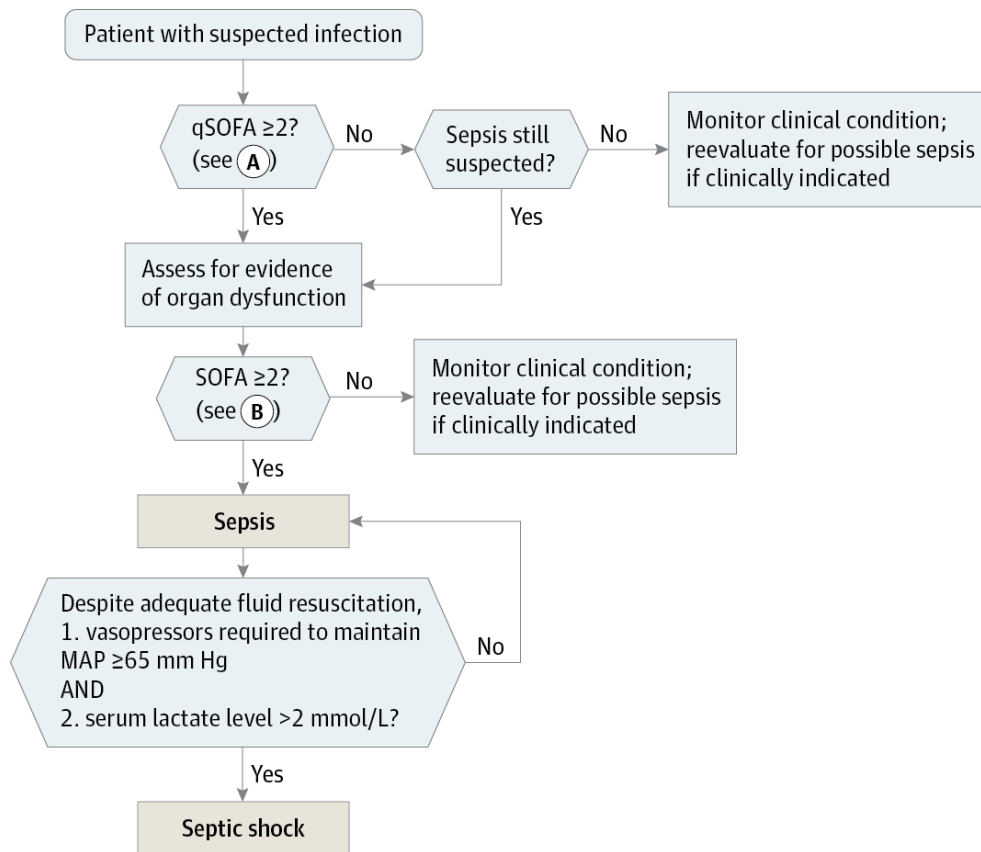


Figure 1: Sepsis-3 definition of sepsis and septic shock. SOFA: Sequential Organ Failure Assessment score, qSOFA indicates quick SOFA; MAP, mean arterial pressure. Reproduced from (Singer M et al., 2016).

Sepsis epidemiology

Sepsis affects as many as 25 million people annually worldwide, is a leading cause of death and one of the most expensive conditions treated in hospitals (Fleischmann et al., 2015; Gotts & Matthay, 2016; Torio & Andrews, 2013). In the USA, treating sepsis may account for up to 40% of all ICU expenses (Torio & Andrews, 2013). It has been estimated that sepsis could cost more than \$27 billion per year to US taxpayers and up to £15 billion yearly in direct and indirect costs to the UK economy (Hex, Retzler, Bartlett, & Arber, 2017; Torio & Andrews, 2013; J.-L. Vincent et al., 2006).

The main risk factors for sepsis include male gender, non-white ethnicity, advancing age, immunosuppression, cancer, genetic factors such as polymorphisms in Toll-like receptor 1 and 4, alcohol consumption, smoking, and vitamin D deficiency (Gotts & Matthay, 2016).

There is wide variation in national-level estimates of sepsis, although most authors reported increasing sepsis rates and decreasing trends in case mortality overall in the last 4 decades (Bouza, López-Cuadrado, Saz-Parkinson, & Amate-Blanco, 2014; Dombrovskiy, Martin, Sunderram, & Paz, 2007; T. E. S. Group, 2004; Kaukonen K, Bailey M, Suzuki S, Pilcher D, & Bellomo R, 2014; Martin, Mannino, Eaton, & Moss, 2003; Stevenson, Rubenstein, Radin, Wiener, & Walkey, 2014). Undoubtedly, sepsis is being more and more recognised and diagnosed, thanks to more global awareness and various education campaigns, which may have contributed to this trend. However, research using standardised diagnostic methods also confirmed the existence of a rising trend. For example, Stevenson et al. calculated standardized mortality ratios from the observed 28-day mortality of usual care participants in clinical trials, for the period 1991 to 2009 (Figure 2) (Stevenson et al., 2014). At the time of this writing, the most recent data available for the United Kingdom had been published by Shankar-Hari et al. (Shankar-Hari, Harrison, Rubinfeld, Rowan, & Myles, 2017). In accordance with the rest of the literature, these authors reported an increase in population incidence and a decrease in hospital mortality over a 5-year period.

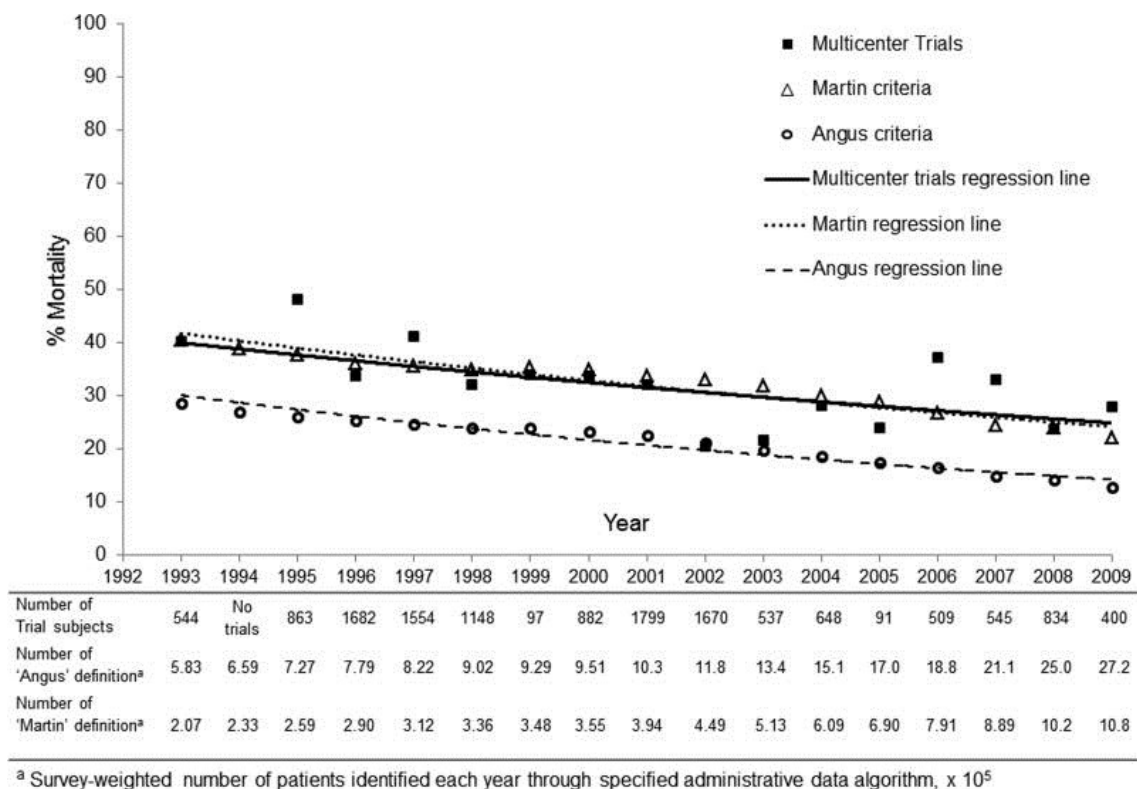


Figure 2: Standardized mortality ratios from the observed 28-day mortality of usual care participants in clinical trials, for the period 1991 to 2009. Reproduced with permission from Stevenson et al. (Stevenson et al., 2014).

Despite improvements in the last decades, mortality remains unacceptably high, at between 15 to 50% depending on the definitions used and the various cohorts (Kaukonen K et al., 2014; P. E. Marik, 2015; Mayr, Yende, & Angus, 2014; Rhee et al., 2017). Altogether, sepsis may claim the lives of around 150,000 and 215,000 people annually, in Europe and the US, respectively (Angus et al., 2001; Artero, Zaragoza, & Miguel, 2012; J.-L. Vincent et al., 2006).

Sepsis physiopathology

Sepsis is caused by bacteria invading the body. These pathogens, when detected by the immune system lead to a massive release, locally and systemically, of cytokines (such as interleukin 1, tumour necrosis factor alpha, etc.), and other inflammatory mediators (Gotts & Matthay, 2016). These mediators are responsible for massive vasodilation, endothelial damage, and local activation of coagulation pathways, increased capillary permeability and decreased systemic vascular resistance. Myocardial dysfunction may also occur and worsen circulation issues.

Altogether, these insults explain that a key clinical feature of patients in the early phase of sepsis is relative (via systemic vasodilation) or absolute (via vascular leakage and reduced fluid intake) hypovolaemia (inappropriate blood volume). Hypovolaemia is a serious condition, especially in frail patients, and can manifest itself through a wide range of features including tachycardia, hypotension, metabolic acidosis, kidney failure, perturbed clotting, respiratory distress or altered consciousness (Gotts & Matthay, 2016; Rhodes et al., 2017). If left untreated, sepsis and hypovolaemia lead to rapid death. Figure 3 shows an example of the different organ failures that can occur in a critically ill patient with septic shock from pneumococcal pneumonia. It should be remembered that pneumonia was by far the leading cause of death in the world before the discovery of antibiotics (Dowling, 1972).

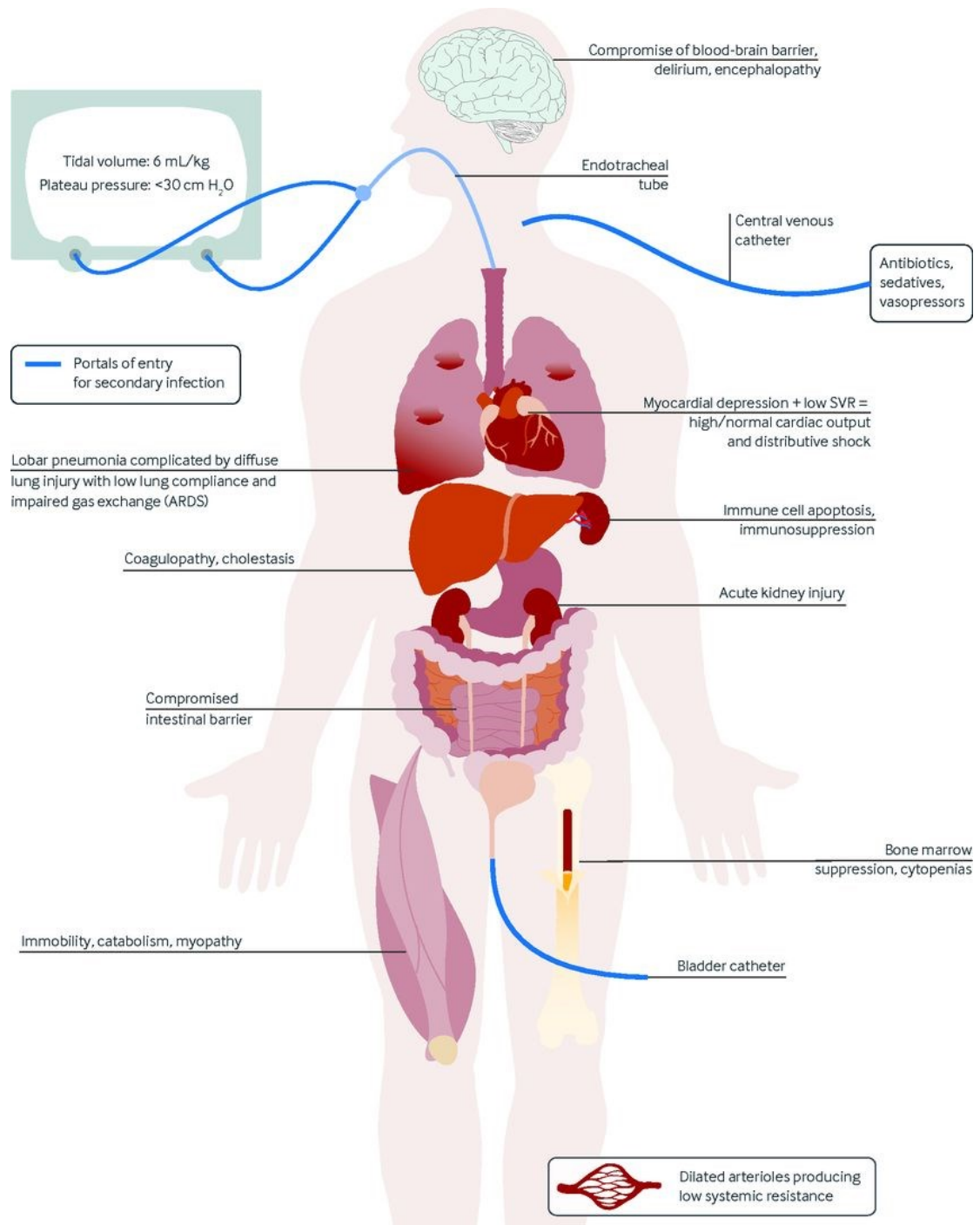


Figure 3: Example of organ failure occurring in a critically ill patient with septic shock from pneumococcal pneumonia. Reproduced with permission from Gotts and Matthay (Gotts & Matthay, 2016).

Current state-of-the-art sepsis management

Several important international endeavours have attempted over the years to improve recognition and treatment of sepsis, noticeably the Surviving Sepsis Campaign (SSC) guidelines, the World Sepsis Day, the Global Sepsis Alliance, among others. The SSC guidelines were introduced in 2002 to provide clinicians with the best available evidence to guide their management of sepsis in an effort to improve patient outcomes (Levy, Evans, & Rhodes, 2018; Rhodes et al., 2017).

The challenges in the management of sepsis include early identification, severity prognostication and providing optimal targeted therapy. The management of a patient with sepsis can be summarised in 3 points (Gotts & Matthay, 2016):

1. Rapid control of the source of infection and treatment with antibiotics
2. Correction of the relative and absolute hypovolaemia with intravenous fluids and/or vasopressors;
3. Treatment of sepsis-induced secondary organ dysfunctions.

Among these three topics, the management of intravenous fluids and vasopressors to correct hypovolaemia remains a central difficulty in sepsis management, and a top research priority (Avni et al., 2015; Byrne & Haren, 2017; Cohen et al., 2015; Gotts & Matthay, 2016; P. Marik & Bellomo, 2016). The classic physiologic rationale for correcting hypovolaemia in sepsis is to restore intravascular volume, cardiac output, and oxygen delivery (Gotts & Matthay, 2016; Rhodes et al., 2017; Semler & Rice, 2016). We have defined hypovolaemia in sepsis as being both absolute and relative, which schematically (and simplistically) can be corrected by the administration of intravenous fluids and vasopressors, respectively.

Intravenous fluids represent sterile solutions which contain water combined with electrolytes or larger molecules and are given intravenously to expand the extracellular volume (“volume expanders”). They are schematically classified in 3 categories: crystalloids, colloids and blood products (Lewis et al., 2018; P. Marik & Bellomo, 2016; Severs, Hoorn, & Rookmaaker, 2015). Crystalloids are solutions of ions which are freely permeable through capillary membranes. They increase plasma volume by about 200 mL for every 1,000 mL given. Normal saline (0.9% sodium chloride) is the most commonly used crystalloid globally. It is isotonic to extracellular fluid but contains a chloride concentration significantly higher than plasma (154 mmol/L), which can lead to hyperchloraemic acidaemia when large volumes are infused. In contrast, so-called balanced crystalloids derived from Hartmann's and Ringer's solutions provide anions that more closely approximate plasma composition.

Colloids are suspensions of large molecules in a carrier fluid with high enough molecular weight to prevent crossing healthy capillary membranes. Colloids can be synthetic (hydroxyethyl starches, gelatins, dextrans) or derivatives of human plasma (albumin solutions). They are thought to be more effective plasma expanders than crystalloids by remaining in the intravascular space and maintaining oncotic pressure. Blood products (red blood cell packs, fresh frozen plasma, pools of platelets, etc.) are not recommended in routine sepsis resuscitation for the correction of hypovolaemia.

Since the advent of intravenous fluids, there has been debate as to which product is the best for patients critically ill from infection. The ideal sepsis resuscitation fluid would increase intravascular volume without accumulating in tissues, contain a chemical composition similar to plasma, and improve patient outcomes in a cost-effective manner (Gotts & Matthay, 2016; Semler & Rice, 2016). No such fluid currently exists. Importantly, capillary leakage is not corrected by fluids or vasopressors, so any type of fluid administration may ultimately exacerbate interstitial oedema and impair organ perfusion. A 2018 Cochrane systematic review of the literature concluded that using colloids versus crystalloids for fluid resuscitation in critically ill people probably makes little to no difference to mortality, but that starches probably slightly increase the need for renal replacement therapy and blood transfusion (Lewis et al., 2018). Crystalloids remain the first line treatment recommended by the latest iteration of the SSC guidelines (Rhodes et al., 2017).

The other drug category of interest is represented by vasopressors. They are powerful drugs that are given by slow intravenous infusions (in general using a central venous access) mainly for their vasoconstrictive properties. They can be classified as adrenergic (norepinephrine, phenylephrine, epinephrine, ephedrine, dopamine) and non-adrenergic (vasopressin and analogues, angiotensin II, nitric oxide synthase inhibitors) (Avni et al., 2015; Russell et al., 2008). In addition to their vasoconstrictive effect, they exert various actions on the cardiovascular system and others (such as positive inotropic effect), which may prompt physicians to use one or another. For example, adrenaline has a potent positive inotropic effect (through beta-1 adrenergic receptor stimulation), and may be administered for patients combining vasoplegia and heart failure. Multiple comparative studies showed that noradrenaline was associated with higher survival than dopamine, and it remains the most commonly administered vasopressor in sepsis (Avni et al., 2015; Rhodes et al., 2017).

The current consensus among intensivists is to titrate fluid administration and vasopressors to reach certain resuscitation targets such as mentation, urine output, mean arterial pressure, central venous pressure, fluid responsiveness, blood lactate, mixed venous oxygen saturation or others, in an attempt to normalise or optimise tissue perfusion (Avni et al., 2015; Gotts & Matthay, 2016; P. Marik & Bellomo, 2016; Rhodes et al., 2017). In this vision, haemodynamic outcomes are used as surrogate markers for survival, the ultimate goal of treatment (Avni et al., 2015). The recommended approach is

to administer intravenous fluids until fluid responsiveness is corrected, and then to initiate vasopressors if blood pressure targets are not achieved. Fluid responsiveness is estimated using various haemodynamic parameters that may be static (central venous pressure, inferior or superior vena cava diameter, pulmonary artery occlusion pressure, etc.) or dynamic (pulse pressure variation, stroke volume variation, inferior or superior vena cava collapsibility/distensibility, plethysmographic variability index, etc.) (Gotts & Matthay, 2016; Mackenzie & Noble, 2014; Rhodes et al., 2017).

Repletion of adequate intravascular volume with intravenous fluids is crucial prior to the initiation of vasopressors since they may be ineffective in the setting of coexistent hypovolaemia. The SSC recommends initiating vasopressors within the first hour in patients who remain hypotensive during or after initial fluid resuscitation (Levy et al., 2018). If maximal doses of a first agent are inadequate, then a second drug should be added (Gotts & Matthay, 2016; Levy et al., 2018). A difficulty comes from the fact that resuscitation parameters are numerous and that optimal targets for each of them are unknown at the patient level and likely dynamic in time. For example, it is still unclear which level of mean arterial pressure should be targeted in sepsis (Asfar et al., 2014; Beloncle, Lerolle, Radermacher, & Asfar, 2013; Corrêa, Jakob, & Takala, 2015). The multiple resuscitation targets and the methods for assessing fluid responsiveness that have been proposed over the years can be seen as yet another indicator of the lack of consensus and definite knowledge about the right approach, which leads to huge practice variation from bewildered clinicians. Additionally, although many clinicians advocate titrating therapies in individuals based on physiological response, we know that this has limitations that short-term physiological improvement doesn't always result in longer-term clinical benefit i.e. survival for patients.

Until recently, the gold standard for sepsis treatment was represented by the Early-Goal Directed Therapy (EGDT) protocols, a highly aggressive and structured approach to sepsis resuscitation (Rivers et al., 2001). It was commonly accepted that septic patients urgently required large amounts of intravenous fluids in order to reverse refractory sepsis-induced tissue hypoperfusion or organ dysfunction. For example, the latest two iterations of the SSC guidelines recommend the administration of 30 ml/kg of crystalloid in the first 3 hours for septic patients with suspected hypovolaemia or initial blood lactate concentration > 4 mmol/L (with no exception) (Dellinger et al., 2013; Rhodes et al., 2017). These guidelines have been implemented and enforced (including by law, for example with Rory's Regulations in the USA) for many years, noticeably in the form of sepsis bundles such as the "sepsis six" which have dramatically impacted routine practice. As a consequence, clinically significant fluid overload and positive fluid balance is a common finding in septic cohorts. For example, a retrospective study measured a positive fluid balance in excess of 6 litres at 24h in 43% of patients with sepsis (Boyd, Forbes, Nakada, Walley, & Russell, 2011).

The harmful effects of administering excessive amounts of fluids and of a sustained positive fluid balance in sepsis are well documented (Acheampong & Vincent, 2015; Byrne & Haren, 2017; Gotts & Matthay, 2016; Kelm et al., 2015; P. E. Marik, 2015). Various studies from multiple countries have established an association between large resuscitation volumes and/or a positive fluid balance in sepsis and acute kidney injury, heart and respiratory failure, hospital, 28-day and 90-day mortality (see example on Figure 4), and the need for medical interventions (diuretics, thoracocentesis, renal replacement therapy for fluid removal, etc.) (Acheampong & Vincent, 2015; Angus et al., 2015a; Boyd et al., 2011; de Oliveira et al., 2015; T. A. Investigators & Group, 2014; T. P. Investigators, 2014; Malbrain et al., 2014; Micek et al., 2013; Mouncey et al., 2015; Rosner et al., 2014; Sirvent, Ferri, Baró, Murcia, & Lorenzo, 2015). Several large randomized controlled trials (RCTs) (including ProCESS, ARISE, and ProMISe) have invalidated EGDT, leaving physicians with little guidance on how to best treat patients (Allen-Dicker, 2015; Angus et al., 2015a; P. E. Marik, 2015; The PRISM Investigators, 2017). Conversely, it was established as early as 20 years ago that a net daily negative fluid balance was closely associated with a reduction in mortality in sepsis (Alsous, Khamiees, DeGirolamo, Amoateng-Adjepong, & Manthous, 2000). This effect seems to persist after adjusting for potential confounders such as patient illness severity.

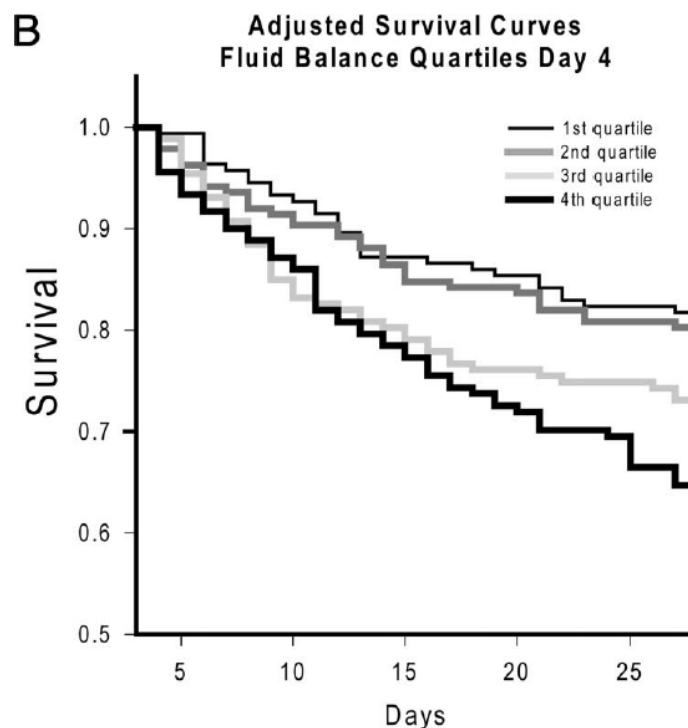


Figure 4: Association between cumulated fluid balance on day 4 post sepsis and 28-day mortality. Reproduced with permission from Boyd (Boyd et al., 2011).

An alternative resuscitation strategy is to use vasopressor therapy earlier (Byrne & Haren, 2017; P. E. Marik, 2015; Waechter et al., 2014) and the potential benefit of this approach in sepsis has been described (Figure 5) (Bai et al., 2014; Beck et al., 2014; Subramanian et al., 2008; Waechter et al., 2014). Noradrenaline has many desirable properties which address several of the physiologic derangements in sepsis. These properties include arterial constriction leading to an increase in blood pressure, positive inotropy and venoconstriction which increases preload and can improve cardiac output and renal perfusion (Bai et al., 2014; Kipnis & Vallet, 2010; Paul E. Marik, 2014). In hypovolaemic patients, noradrenaline may improve preload while intravenous fluids are simultaneously being infused.

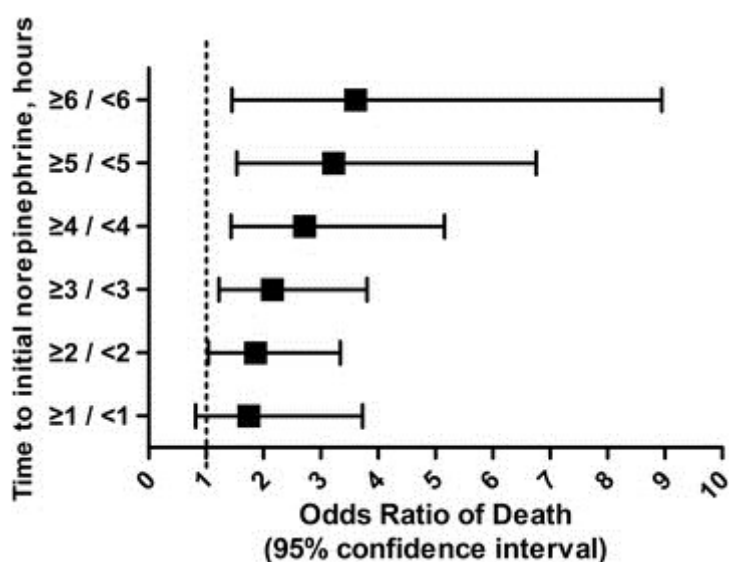


Figure 5: Association between norepinephrine administration delay (from the onset of septic shock) and hospital mortality. Reproduced from Bai (Bai et al., 2014).

To sum up, no tool is currently available to individualise the treatment of sepsis, while a more personalised medicine has been hoped for (Byrne & Haren, 2017; Dellinger et al., 2013; P. Marik & Bellomo, 2016; P. E. Marik, 2015; J.-L. Vincent, 2016). Several key clinical questions in sepsis management remain unanswered, such as the right balance between fluid and vasopressors, the timing for initiating those drugs, the rate of fluid administration, the correct volume of intravenous fluid to administer during initial resuscitation and later stages, resuscitation targets and fluid balance targets (Acheampong & Vincent, 2015; Asfar et al., 2014; Beck et al., 2014; Beloncle et al., 2013; Corrêa et

al., 2015; de Oliveira et al., 2015; Gotts & Matthay, 2016; Malbrain et al., 2014). The cut-off point between giving too little and too much fluid is dynamic and difficult to assess, and both conditions lead to adverse outcomes (Figure 6). Clinicians have difficulty identifying which patients need fluid: several studies have shown that approximately 50% of fluid boluses fail to achieve an increase in cardiac output, meaning that they were given inappropriately (Mackenzie & Noble, 2014).

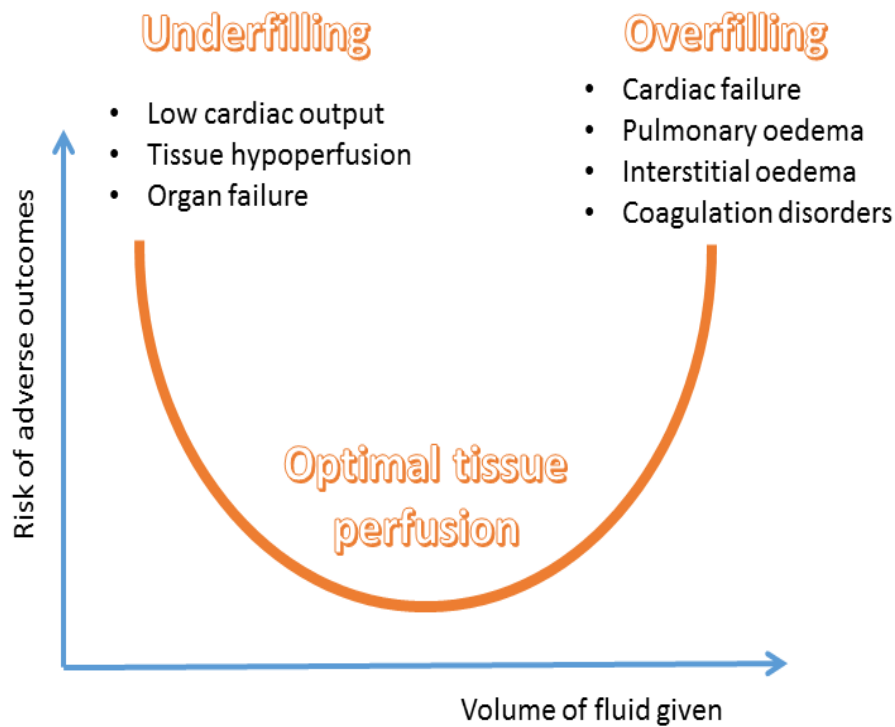


Figure 6: Fluid load in sepsis versus risk of organ failure and other complications. Modified from Bellamy (Bellamy, 2006).

As a consequence of the uncertainty around sepsis resuscitation, clinical variability in treatment is extreme, with consistent evidence that suboptimal decisions lead to poorer outcomes. It has been suggested that current research paradigms (in particular in the form of RCTs) have stalled, and that new approaches are needed in order to achieve further improvements in outcomes (Bosurgi, 2015; Leo Anthony Celi, Zimolzak, & Stone, 2014; Cohen et al., 2015; M. Ghassemi, Celi, & Stone, 2015).

RCTs remain the gold standard for clinical knowledge discovery (M. Ghassemi et al., 2015; Murad, Asi, Alsawas, & Alahdab, 2016). The reality is that most treatment comparisons have never been tested by an RCT and that only a small proportion of medical decisions are based on RCT-supported evidence (M. Ghassemi et al., 2015). The British Medical Journal analysed the harms and benefits of 3,000 medical interventions in RCTs and came to striking conclusions (Smith, Street, Volk, & Fordis, 2013).

Half of the interventions were of unknown effectiveness, while only about a third of the treatments were shown to be beneficial (11%) or likely to be beneficial (23%). Exponential combinations of patients, diseases and treatments can never be exhaustively tested by RCTs. RCTs often include highly selected populations, hence their conclusions are not generalizable to other groups, such as those of different age, ethnic origins or with particular comorbidities or medications. Often, conclusions from RCTs are simply not applicable to real-world decisions about real-world patients (M. Ghassemi et al., 2015)!

Large-scale retrospective health data analytics, combined with biostatistics and/or machine learning represent an opportunity to address clinically relevant questions using another approach than RCTs (Leo A. Celi, Csete, & Stone, 2014; Leo A. Celi, Mark, Stone, & Montgomery, 2013; M. Ghassemi et al., 2015). Retrospective studies, in general, are useful for testing the association between multiple exposures and outcomes and often can be conducted faster than RCTs and for less money (Frieden, 2017; Murad et al., 2016). The risk of bias and confounding effects are well documented.

The potential applications for health data analytics in acute care environments include predictive models for prognostication and early alerting, reporting analytics of patient stays, triage, readmission, adverse events, clinical decompensation, and optimization of treatment decisions and care pathways (M. Ghassemi et al., 2015). Over the last few years, ICU databases have allowed to develop predictive models with actionable outputs that potentially influenced clinical practice and led to quantifiable improvements in process and/or outcome (Leo A. Celi et al., 2014, 2013; Leo Anthony Celi, Hinske, Alterovitz, & Szolovits, 2008; M. Ghassemi et al., 2015; M. M. Ghassemi et al., 2014).

The opportunity for machine learning in decision-making in healthcare

Machine learning is an application of artificial intelligence (AI), where a machine analyses data and acquires the ability to execute complex (“smart”) tasks without being explicitly programmed. Machine learning has been suggested as a novel approach to assist decision-making in healthcare (Bennett & Hauser, 2013; Murdoch TB & Detsky AS, 2013). This concept fits into the vision of a data-driven healthcare system, where the previous medical cases and information are converted into clinical tools and decision support systems that are applied to new patients, whose information is thereafter added to the global database, leading to further increases in knowledge (Leo Anthony Celi et al., 2014).

Reinforcement learning (RL) is a category of machine learning tools where a virtual agent learns from trial-and-error an optimized set of rules – a policy – that maximizes an expected return (Bennett & Hauser, 2013; Sutton & Barto, 2018). Similarly, a clinician’s goal is to make therapeutic decisions in order to maximize a patient’s probability of a good outcome (Bennett & Hauser, 2013; Schaefer, Bailey,

Shechter, & Roberts, 2005). RL has many desirable properties that may help medical decision-making. Their intrinsic design that uses sparse reward signals makes them well suited to overcome the complexity related to the heterogeneity of patient responses to medical interventions and delayed indications of the efficacy of treatments (Sutton & Barto, 2018). Importantly, these algorithms can infer optimal strategies from suboptimal training examples, which is what human clinicians' decisions provide. Clinical databases present an opportunity to study medical questions where practice variation exists, as a result of either lack of or conflicting medical knowledge (Leo A. Celi et al., 2014, 2013; M. Ghassemi et al., 2015). RL can take advantage of this variability in clinical practice, and identify which decision(s) appear(s) to be the most optimal for a given group of patients or even at the individual patient level.

Introduction to reinforcement learning and notation

As defined by Sutton and Barto: “Reinforcement learning is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal” (Sutton & Barto, 2018). RL is a class of machine learning algorithms, whose goal is to estimate an optimal set of rules (a policy) that maximises some form of reward or return, using a model of a decision process. *“The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics—trial-and-error search and delayed reward—are the two most important distinguishing features of reinforcement learning.”* (Sutton & Barto, 2018).

Various Markov model frameworks can be used to deploy RL algorithms, the simplest being a discrete Markov decision process (MDP) (Puterman, 1994; Sutton & Barto, 2018). In an MDP, we model the interactions of an agent and an environment, in which the agent follows a policy π and acts on the environment. The environment reacts by transitioning into a new state and releasing a reward if the new state has more desirable properties than the previous state (Figure 7). When this happens, the decision that led to this transition is being reinforced – hence the term RL.

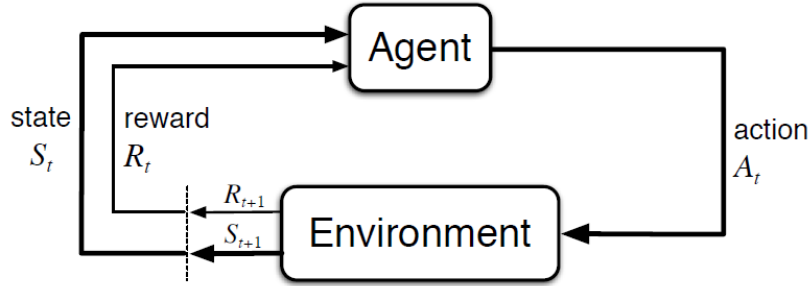


Figure 7: The agent-environment interaction in an MDP. Reproduced from Sutton and Barto (Sutton & Barto, 2018).

In medical applications, the true patient physiological state is only partially represented by the data available, and therefore the disease process could be formulated as a Partially Observable MDP (POMDP). A POMDP accounts for the uncertainties in the decision maker’s observations of the actual state of the environment (Spaan, 2012). In our problem, we observe patient parameters (consciousness, heart rate, blood pressure, arterial pH...) which depend on the actual health state of a patient, but only represent a part of the information about the true state. A POMDP extends the MDP framework by adding observations (a finite set of observations of the state) and an observation function (which captures the relationship between the observations and the state, and can be action dependent). These observations can be used to model perceptual aliasing (the fact that many states can give the same observation), noisy or faulty sensors (e.g. electrode disconnection for heart rate measurement), or both (Spaan, 2012). Since solving a POMDP is computationally expensive and require additional assumptions, we simplified the POMDP into an MDP to approximate patient trajectory and to model the decision-making process (Bennett & Hauser, 2013; Puterman, 1994; Sutton & Barto, 2018). The MDP model naturally captures the variability in physiological responses to clinical events as well as the variability in patients’ trajectories. Markov models deal very well with time series since they are able to capture the dependencies between variables, but also the serial correlation in the measurements (Aghabozorgi, Seyed Shirshorshidi, & Ying Wah, 2015). In our model, the patient represents the environment in which the physician (the agent) acts (Figure 8 and Figure 9).

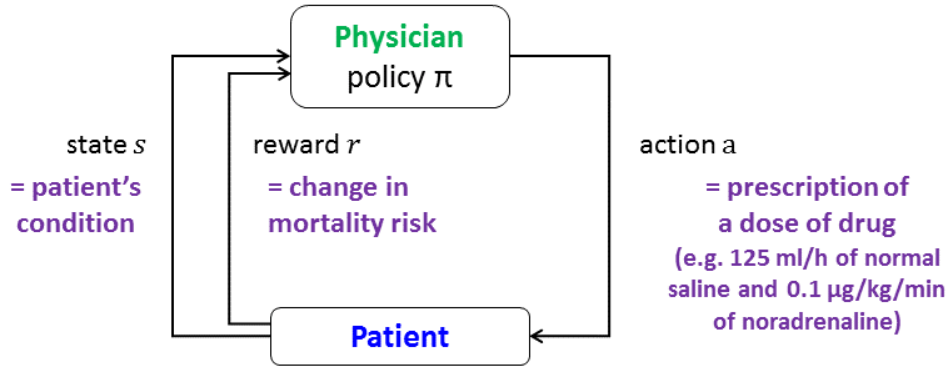


Figure 8: The framework of RL applied to the healthcare setting, showing the physician-patient interaction. Adapted from (Sutton & Barto, 2018). The policy π is the set of rules controlling which action is taken while in a particular state.

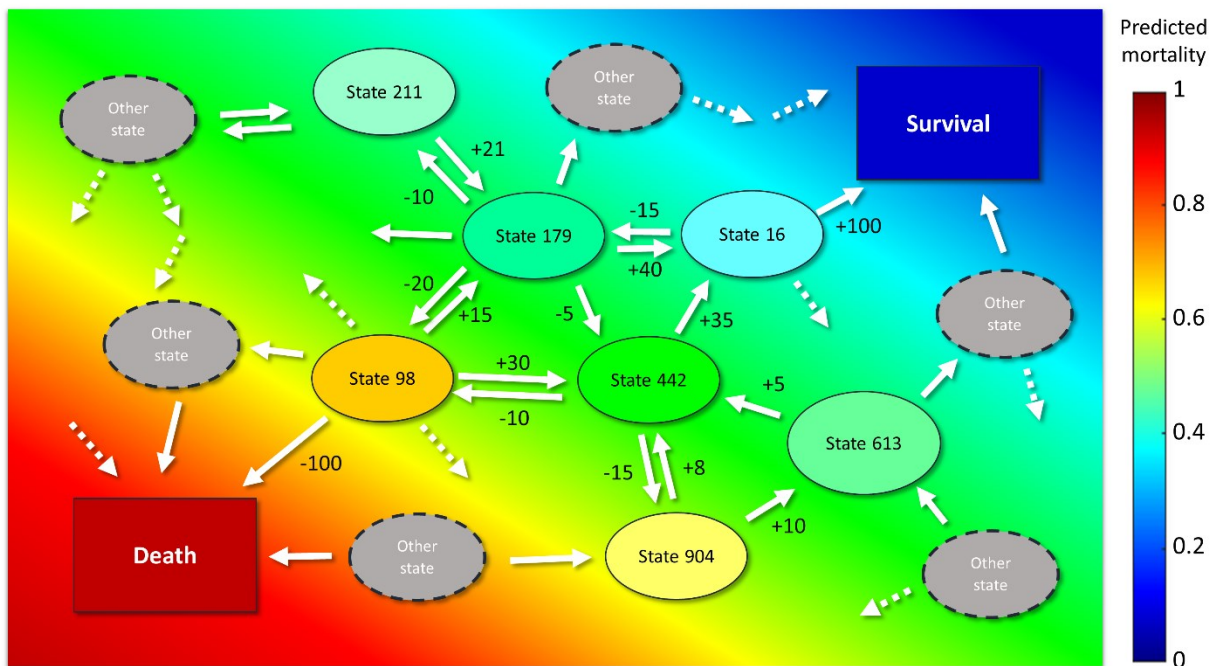


Figure 9: Conceptual representation of trajectories of critically ill patients through different states, from admission to survival or death. Each known possible transition is associated with a reward or penalty (the value of the action), ranging from -100 to +100 points in this example, depending on how the treatment administered probabilistically affects the patient prognosis. In each state, the action of highest value is the most optimal action.

Formally, a Markov decision process (MDP) is defined as a tuple $\{S, A, T, r, \gamma\}$ where S is a set of states, A is a set of actions, $T : S \times S \times A \rightarrow [0, 1]$ is a probability mass function defining a distribution over next states S' for each state and action, $r : S \times A \rightarrow [-100, 100]$ is a bounded reward function, $\gamma \in [0, 1]$ is a discount factor, which can be thought of as an interest rate (an immediate reward is worth more than a reward in the future). Our notation assumes that the state, action, and reward sets are finite. An agent samples actions from a policy, $\pi : S \times A \rightarrow [0, 1]$, which is a probability mass function on A conditioned on the current state. A policy is deterministic if $\pi(a|s) = 1$ for only one a in each s , or probabilistic if $\pi(a|s) = p(a|s, \pi)$. A trajectory, H of length L is a state-action history, $S_0, A_0, \dots, S_{L-1}, A_{L-1}$ where $S_0 \sim d_0$, d_0 is a probability mass function over initial states, $A_t \sim \pi(\cdot|S_t)$, and $S_{t+1} \sim T(\cdot|S_t, A_t)$. The return of a trajectory H is $g(H)$. The policy, π , and transition dynamics, T , induce a distribution over trajectories, p_π . We write $H \sim \pi$ to denote a trajectory sampled by executing π (i.e., sampled from p_π). The expected discounted return of a policy π is defined as $V(\pi) := E[g(H)|H \sim \pi]$. Two separate policies are defined: π_b the behaviour policy, the policy followed by clinicians, and π_e , the evaluation policy, AI policy or “optimal” policy for the model. The training ICU dataset provides a set of n historical trajectories $D = \{H_1, \dots, H_n\}$, where $H_i \sim \pi_b$.

Next, we must select and justify the choice of RL algorithms in order to 1) evaluate π_b and 2) generate π_e .

Reinforcement learning algorithms

Clinicians’ policy evaluation

We performed an evaluation of π_b , the policy of clinicians using temporal difference learning (TD-learning) of the Q function, by observing all the drug prescriptions in existing records and computing the average value of each treatment option, at the state level (Sutton & Barto, 2018). Because the algorithm uses existing episodes, it is said to rely on offline sampling.

The advantage of TD-learning versus policy iteration is that it does not require knowledge of the MDP (model-free), and makes it possible to learn simply from sample trajectories (Sutton & Barto, 2018). It was computed from actual patient episodes of successive state-action pairs, with resampling, using the following iterative procedure (Algorithm 1). A learning rate α of 0.1 was selected.

- 1- Initialise the Q value as a zeros array
- 2- Pick one episode randomly from the training data
- 3- Compute the Q update formula for all the time steps of the episode, in reverse order:

$$Q^\pi(s, a) \leftarrow Q^\pi(s, a) + \alpha \cdot (r + \gamma \cdot Q^\pi(s', a') - Q^\pi(s, a)) \quad (1)$$

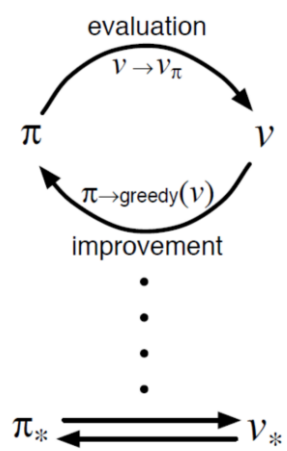
With $Q^\pi(s, a)$ the Q value of the current {state, action} tuple considered, $Q^\pi(s', a')$ the Q value of the next {state, action} tuple, α the learning rate and r the immediate reward.

- 4- Repeat steps 2 and 3 for a total of 500,000 iterations

Algorithm 1: Pseudocode for TD-learning.

Optimal policy estimation

We learned π_e , a theoretical optimal policy for the simplified MDP using in-place policy iteration, which identified the decisions that maximized the long-term sum of rewards, hence the expected survival of patients (Sutton & Barto, 2018). Policy iteration is a type of dynamic programming algorithm that starts with a random policy that is iteratively evaluated then improved until converging to an optimal solution (see schematic below, from (Sutton & Barto, 2018), page 63).



Policy iteration was implemented using the procedure summarised in Algorithm 2.

- 1- Randomly initialise the optimal policy
- 2- Evaluate the value function of the current optimal policy
- 3- Improve the current optimal policy using a greedy policy definition
- 4- Repeat steps 2 and 3 until no further improvement in the optimal policy is seen

Algorithm 2: Pseudocode for policy iteration.

After convergence, the AI policy π^* corresponds to the actions with the highest state-action value in each state:

$$\pi^*(s) \leftarrow \operatorname{argmax}_a Q^{\pi^*}(s, a) \forall s \quad (2)$$

The value V of a policy π was computed using the recursive, self-consistent Bellman equation for V^π and represents the expected return when starting in s and following π thereafter:

$$V^\pi(s) = \sum_a \pi(s, a) \sum_{s'} T(s', s, a) [R(s') + \gamma V^\pi(s')] \quad (3)$$

We also estimated the AI policy using off-policy Monte Carlo control, which led to identical results but was roughly 1,000 times slower to run. In Monte Carlo control, the optimal policy is learnt from exploring the state-action space, generating virtual patient trajectories and using complete sample returns (from any state until discharge or death). The general concept is that the Q function will be learnt from averaging the returns from multiple visits to each state. The convergence of the algorithm is guaranteed by the use of a soft policy definition, which maintains a degree of exploration during the generation of episodes, so the probability to visit any $\{s,a\}$ pair is not null. We used an off-policy variant of the algorithm since the policy used to generate the data (the behaviour policy, or the clinicians' policy) may be unrelated to the policy that is evaluated and improved (the estimation policy, or the AI policy). The estimation policy is updated using a greedy policy definition. Off-policy methods are appropriate to learn from data generated by a non-learning controller or from a human expert (Sutton

& Barto, 2018). In the end, policy iteration was the only method used to estimate the optimal policy because it was computationally more efficient.

Challenges of model evaluation

Retrospective validation of a learning algorithm is challenging because it is impossible to know the outcome of an action that was not taken. We must find reliable methods to assess the value of a policy that was never deployed and is different from the one that was executed by clinicians. Formally, we want to predict the performance of a newly generated RL policy (π_e , the evaluation policy, the AI policy) given historical data that has been generated by a different policy (π_b , the behaviour policy, the clinicians' policy). This problem is called off-policy policy evaluation (OPE).

Why is it so difficult?

Both shallow and deep RL algorithms have been successfully applied to a variety of artificial tasks such as solving randomly generated MDPs (Precup, Sutton, & Singh, 2000), pathfinding (Hanna, Stone, & Niekum, 2016; Jiang & Li, 2015), self-balancing poles, the mountain car problem (Jiang & Li, 2015; P. S. Thomas, Theodorou, & Ghavamzadeh, 2015), playing Atari games (Mnih et al., 2015) even without reward function (Aytar et al., 2018), chess or the game of go (Silver et al., 2016). Why are we not seeing myriads of high impact research applying RL to healthcare problems? The following section on “related work” presents a few examples, most of which have limited value (in particular because of limited state and action spaces) and have made little to no impact to clinical practice. However, the list of candidate questions that could theoretically be addressed with RL in intensive care alone is infinite: blood glucose control, renal replacement therapy and sedation, sequential antibiotic dosing, etc.

We explored some of these arguments in a separate Nature Medicine publication (Gottesman et al., 2019). First of all, a key limitation until recently has been the lack of available data to conduct such research. Digitization of healthcare is now widespread, having reached for example 96% of all non-federal acute care hospitals in the USA in 2015 (Office of the National Coordinator for Health Information Technology, 2016). Next, converting raw electronic health record (EHR) data into an analysable dataset requires a tremendous effort, expertise in data science, healthcare and the individual EHR software, in order to address many issues such as siloed data, anonymisation, harmonization, encoding of clinical concepts, etc. (Leo A. Celi et al., 2014; Leo Anthony Celi et al., 2014). The next difficulty is that healthcare is a high-risk environment: deploying a bad policy would be dangerous

and/or costly (P. S. Thomas & Brunskill, 2016; P. S. Thomas et al., 2015). We need guarantees of the safety of the policy before implementation. In medicine, we only have a limited amount of training data, whereas virtual agents (such as those used in computer simulations) can collect millions of hours of simulated trials (e.g. gameplay). In these other settings, the environment is in general fully specified: at any moment, all the information available to make the best possible move (Figure 10). On the other hand, in medicine, the patient data available represents only a fraction of the data that could be collected, which itself represents an imperfect and incomplete representation of the true patient health state. In some sense, the data available could be compared to “looking through the keyhole of physiology”. In artificial computer science problems, RL agents learn and improve by trial and error. This is impossible in healthcare: we cannot let an RL agent try various (partially random or unrefined) policies on human subjects and simply learn by trial and error. Also, training time would be prohibitive if only one trial could be run in the real world at a time. These challenges are to some extent common when trying to apply RL in other domains such as robotics, where hardware cost and training speed precludes large-scale learning by trial and error (Kormushev, Calinon, & Caldwell, 2013). High-fidelity simulation of the robots in their environment can allow some of these difficulties to be overcome. In medicine, there is no high-fidelity simulator that would allow testing the policy without putting human lives at risk.

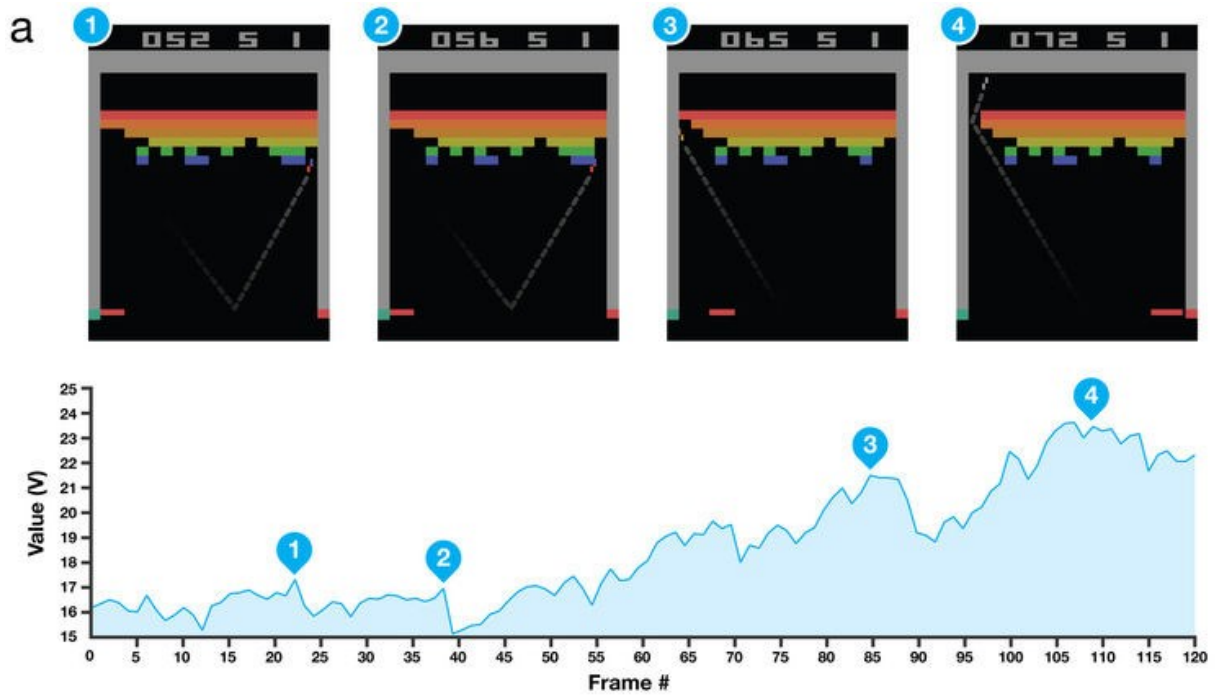


Figure 10: Breakthrough 2015 Nature paper on deep RL applied to learning to play classic Atari games “pixel to action”. Here the environment is fully specified: at any moment, all the information is available to make the best possible decision, which is not true for medical problems. Reproduced from (Mnih et al., 2015).

Formalising off-policy evaluation

In OPE tasks, we want to predict the performance of π_e the evaluation policy (generated by the RL agent) given historical data that was generated by π_b the behaviour policy (the clinicians in our case).

Schematically, OPE can be model-based or importance sampling based (Jiang & Li, 2015; P. S. Thomas & Brunskill, 2016). The first approach is model-based, which Jiang & Li call “regression-based”, and Thomas refers to as computing the “approximate model” estimator, where an approximate model of the MDP is constructed using all the available data (fitted to the data via regression), then used to compute the performance of the evaluation policy, which is used as an estimate of $V(\pi_e)$ (P. S. Thomas & Brunskill, 2016). *“Such a regression-based approach has a relatively low variance and works well when the model can be learned to satisfactory accuracy. However, for complex real-world problems, it is often hard to specify a function class in regression that is efficiently learnable with limited data while at the same time has a small approximation error. Furthermore, it is in general impossible to estimate the approximation error of a function class, resulting in a bias that cannot be easily quantified.”* (Jiang

& Li, 2015). “The second class of approaches are based on the idea of importance sampling (IS), which corrects the mismatch between the distributions induced by the target policy and by the behavior policy. Such approaches have the salient properties of being unbiased and independent of the size of the problem’s state space, but its variance can be too large for the method to be useful when the horizon is long” (Jiang & Li, 2015).

An important motivation for off-policy evaluation is to guarantee safety before deploying a policy. For this purpose, we have to characterize the uncertainty in our estimates, usually in terms of a confidence interval (CI), which led to the concept of high-confidence off-policy evaluation (HCOPE) (Gottesman et al., 2019; Hanna et al., 2016; Jiang & Li, 2015; P. S. Thomas & Brunskill, 2016; P. S. Thomas et al., 2015). Different approaches can be used to achieve this, but the most data-efficient appears to be bootstrapping (more detail below in chapter 4) (Hanna et al., 2016; P. S. Thomas & Brunskill, 2016; P. S. Thomas et al., 2015).

Related work

MDPs are a powerful and appropriate technique for modelling medical decision (Schaefer et al., 2005). They are most useful to formulate problems involving sequential, stochastic and dynamic decisions like medical treatment, for which they can find optimal solutions (Schaefer et al., 2005). Despite this tremendous potential, there have been very few applications of RL in healthcare, for the reasons discussed above (Bennett & Hauser, 2013).

Early models such as the one proposed by Tsoukias et al., that dealt with antibiotic treatments options in intensive care (Tsoukalas, Albertson, & Tagkopoulos, 2015) were limited by a restricted state space (only 10 states, defined by expert opinion) and set of features (only 6). Other early applications of Markov models in healthcare explored various problems such as insulin therapy in diabetes (Bothe et al., 2013; Daskalaki, Diem, & Mougiakakou, 2016), propofol anaesthesia (Moore et al., 2014), liver transplant (Alagoz, Maillart, Schaefer, & Roberts, 2004), HIV therapy (Shechter, Bailey, Schaefer, & Roberts, 2008), breast cancer (Maillart, Ivy, Ransom, & Diehl, 2008), Hepatitis C progression (Daniel M Faissol, 2007), statin therapy (Denton, Kurt, Shah, Bryant, & Smith, 2009) and hospital discharge management (Kreke, 2007).

Lately, a few exciting projects made use of the high dimensionality of patient data, in ICU (Prasad, Cheng, Chivers, Draugelis, & Engelhardt, 2017), in ophthalmology (T. D. R. Group, 2017) or for HIV therapy (Parbhoo, Bogojeska, Zazzi, Roth, & Doshi-Velez, 2017). Prasad’s used the Medical Information Mart for Intensive Care, version 3 (MIMIC-III) database to build a continuous MDP to

model sedation and mechanical ventilation. The authors derived an optimal strategy and demonstrated that the higher the distance between actual and optimal policies, the worse the outcomes (Figure 11). We use some similar validation methods (see for example Figure 38).

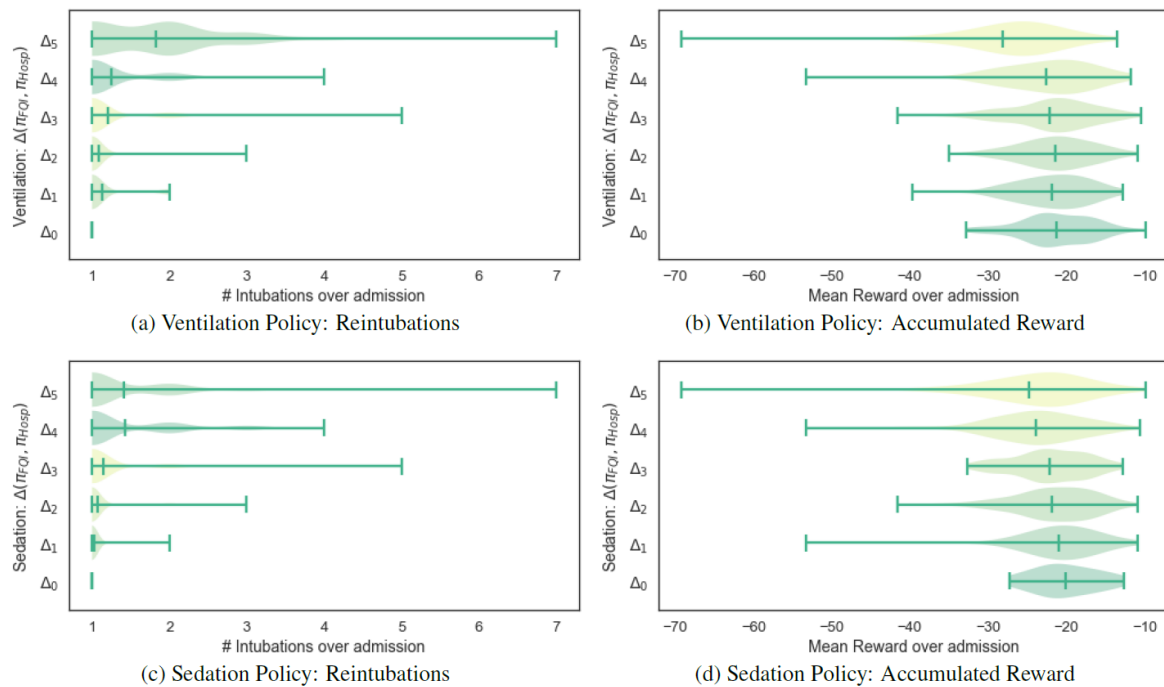


Figure 11: Evaluation method suggested by Prasad et al. Patient outcome is plotted against the distance between the AI and the clinicians’ policies, demonstrating that outcomes are better when the actual actions match the policy generated by the RL agent. Reproduced from (Prasad et al., 2017).

RL was also used to model the best interval for monitoring of diabetic retinopathy (T. D. R. Group, 2017). Parbhoo’s work combined an RL model to kernel-based methods (so-called mixture-of-experts approach) to derive an optimal management strategy for HIV therapy (Parbhoo et al., 2017).

Hypotheses and objectives

I hypothesised that RL could be used to model the dynamics of adult patients with sepsis and help identify optimal treatment strategies. The objective of this research to build a framework to implement RL algorithms from the secondary analysis of existing ICU medical records, in order to optimise the management of intravenous fluids and vasopressors in adult patients with sepsis.

This hypothesis implies several sequential assumptions:

1. The true patient health state is represented to a sufficient extent by the data available in existing ICU databases;
2. It is possible to retrospectively identify sepsis according to the recognized definition;
3. The trajectories of patients with sepsis can be modelled with an MDP;
4. Fluid and vasopressor therapy (stochastically) influence the transition between health states in this target patient population;
5. The value of drug dosage can be quantified (with a confidence interval) in terms of their impact on the risk of mortality;
6. An optimal policy can be learnt, that optimises patient mortality;
7. Methods exist to quantify the value (with a confidence interval) of this newly generated optimal policy;
8. This optimal policy has a higher value than the clinicians' treatment strategy.

Overview of the approach

The data flow in this research is outlined in Figure 12.

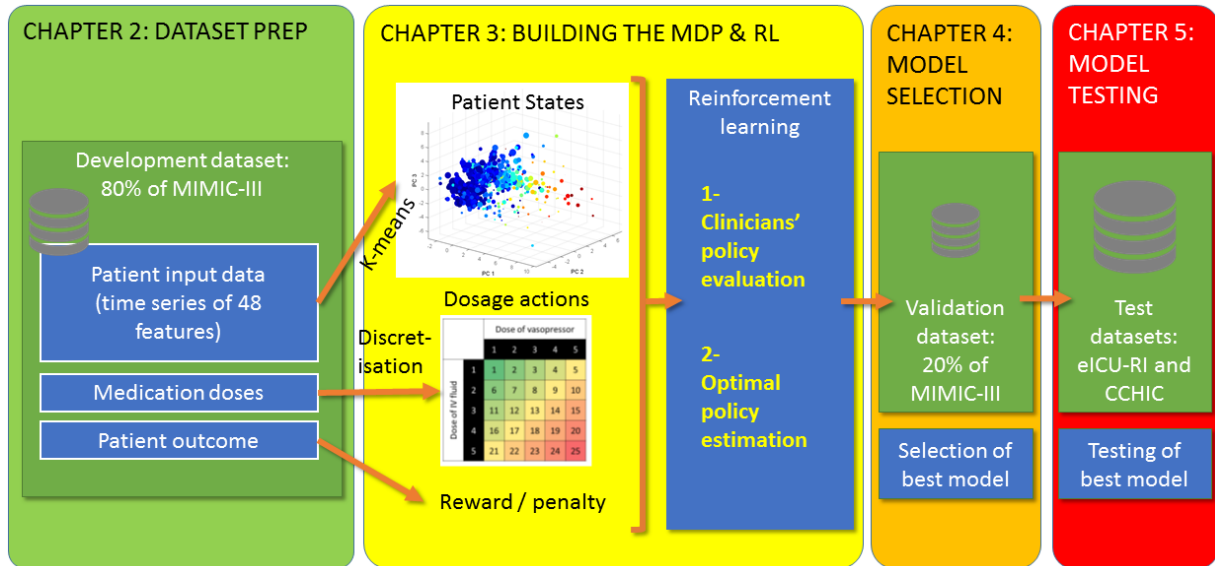


Figure 12: Data flow in the project and content of the thesis chapters. Eighty percent of the MIMIC-III dataset was used to define the elements of the Markov decision process. The dose of intravenous fluids and vasopressors were discretized into 25 possible actions. Patients' survival at 90 days after ICU admission defined rewards. RL was used to evaluate the value of clinicians' policy and estimate optimal treatment strategies. The model was validated on the remaining 20% MIMIC-III data then tested on two independent datasets from the e-Research Institute (eRI) and CCHIC databases.

Chapter 2: Data extraction and pre-processing

This chapter contains the description of how the three separate datasets used in this project were prepared.

The datasets

Three databases were used in this research: MIMIC-III (version 1.4), eRI (version 2.0) and CCHIC (version 1.0). All databases contain high-resolution patient data including demographics, vital signs time series, laboratory tests, illness severity scores, medications and procedures, fluid outputs, clinical notes, and diagnostic coding. CCHIC does not contain information about intravenous fluid intake, so this set of decisions could not be assessed. In all databases, the data of interest can be constant (age, gender) or time-varying, and features can be binary (e.g. readmission), categorical (e.g. gender) or continuous (e.g. blood sodium).

MIMIC-III is an open-access anonymized database of 61,532 distinct admissions between 2001 and 2012 in 6 ICUs (Coronary Care Unit, Cardiac Surgery Recovery Unit, Medical Intensive Care Unit, Surgical Intensive Care Unit, Trauma Surgical Intensive Care Unit and Paediatric Intensive Care Unit) at the Beth Israel Deaconess Medical Center, a large Boston teaching hospital (Goldberger et al., 2000; Alistair E. W. Johnson et al., 2016). The data in MIMIC-III was collected via two different critical care information systems: Philips CareVue Clinical Information System (models M2331A and M1215A; Philips Healthcare, Andover, MA) and iMDsoft MetaVision ICU (iMDsoft, Needham, MA). MIMIC-III is unique among the databases used because it linked hospital electronic health record with the Social Security Administration Death Master File, so mortality even after hospital discharge is known.

The full version of the eRI contains more than 3.3 million admissions in 459 ICUs across the U.S.A, recorded between 2008 and 2016 via the Philips tele-ICU program. This project allows clinicians to remotely monitor and guide the management of multiple ICU patients across multiple sites, from a centralised location. As such, the system streams a wide range of patient information, which is then collected and stored by the company. It contains the same data types as MIMIC-III, but data quality is very heterogeneous across the sites, so processes had to be developed to filter out the ICUs where data quality was insufficient (more on this later in this chapter).

The full MIMIC database is available to researchers who complete a human research ethics training programme and sign a data use agreement. Access to MIMIC has been granted to more than 4,000 individuals and institutions throughout the world. The full version of the eRI is not openly available but

access was granted by the Philips Research Institute research committee in December 2016. A subset of over 200,000 patients from the eRI has been made available to the public in 2017, following the MIMIC model, and is called the eICU Collaborative Research Database (<https://eicu-crd.mit.edu/>).

The Critical Care Health Informatics Collaborative database (CCHIC) comprises data from 33,535 unique patients (42,232 ICU admissions) admitted between 2014 and 2017 into 11 general adult medical and surgical ICUs at the five founding National Institute of Health Research (NIHR) Biomedical Research Centres (BRCs) at Cambridge, Guy's, Kings' and St Thomas', Imperial, Oxford and University College London (UCL) (Harris et al., 2018). The current dataset includes 264 fields comprising 108 constant fields (hospital, unit, patient and episode descriptors, recorded once per admission), and 154 time-varying physiology and therapeutic fields (recorded hourly, daily etc.). Data is currently being added on a quarterly basis, and the full database is hosted in a "safe haven" with restricted physical access within the Wolfson Institute for Biomedical Research, based in the UCL Cruciform building. An anonymised data subset is freely available, but access to the full identifiable dataset is restricted to a selected number of researchers internal to UCL.

All the three datasets are relational SQL databases. MIMIC-III and eRI (PostgreSQL) were accessed and queried using HeidiSQL (version 9.4.0.5125), while CCHIC (SQLite) was accessed from Rstudio (version 1.1.383) using the packages *cleanEHR* (<https://github.com/ropensci/cleanEHR>), *dplyr* and *dbplyr*. With regards to data readiness level, all three databases were "band C" as per Lawrence's taxonomy, meaning that the data was available and ready to be loaded into a data extraction software, but no information was known at this point on data trustworthiness, missingness, etc. (Lawrence, 2017).

The institutional review board (IRB) of the Massachusetts Institute of Technology (No. 0403000206) and Beth Israel Deaconess Medical Center (2001-P-001699/14) approved the use of MIMIC-III for research. The use of the eRI database was approved by the eICU research committee and exempt from IRB approval as the database security schema and the re-identification risk were certified as meeting safe harbour standards by Privacert (Cambridge, MA) (45 Code of Federal Regulations 164.514(b)(1) and Health Insurance Portability and Accountability Act Certification no. 1031219-2). The UK National Research Ethics Service granted an exemption to the common law duty of confidentiality for the CCHIC project (14/LO/103) (Harris et al., 2018). Data sharing agreements were signed between the participating NHS Trusts and UCL which hosts the Data Safe Haven (DSH) where CCHIC is stored. The DSH is certified to the ISO/IEC 27001:2013 information security standard and conforms to the NHS Digital's Information Governance Toolkit. Since this research was a secondary analysis of fully anonymised data and that all databases had been individually certified for research use, individual patient consent was not required.

Data requirements

Sepsis is a heterogeneous syndrome that presents as an infinite number of clinical features and evolutions. If we hope to capture even a fraction of this heterogeneity in our model, we require a high dimensional dataset, comprising many features (synonyms: parameters, variables), each representing a fraction of the information about the patient’s clinical status. To capture the time-varying character of patients’ status, we require the data to be represented as time series, built from data encoded at successive time points. The higher the sampling frequency of the data, the more granular the model will be. Obviously, many patient features are not sampled very frequently, for example, many blood tests, so increasing the granularity of the model amplifies the quantity of missing data.

The quality of the data is a key determinant of model quality and robustness. Processes were developed to test for data quality, in particular to identify hospitals with low data availability in eRI (see below). We also check for the presence of erroneous values, outliers and missing data (see below). When referring to the “readiness level” of the data after the dataset preparation, our modelling requires a band B dataset, meaning that the data quality must be trusted or that missing values, outliers and noise must be quantified (Lawrence, 2017).

The most straightforward way to represent a patient in a computational model is using numerical values. Fortunately, the bulk of the data collected in ICU patients is numerical and structured (heart rate, blood pressure, arterial blood gases, etc.). Limiting patient features to numerical data is a limitation but capturing text-based information (past medical history, diagnoses, clinical signs) is much more complex and requires methods such as natural language processing to produce word embeddings (see for example *word2vec* (Mikolov, Chen, Corrado, & Dean, 2013)). Novel approaches based on deep-learning to generate patient representations have been proposed (M. Ghassemi et al., 2014; Miotto, Li, Kidd, & Dudley, 2016; Rajkomar et al., 2018). Also, this information is often sampled at low frequency, at best a few times per day. Finally, representing this information in a (numerical) database means in general transforming text concepts in a potentially very vast number of binary or categorical variables, while making trade-offs in the translating process (e.g. keeping only one feature for “cardiac history” when it actually represents a large variety of diagnoses). A final limitation to text-based medical data is variability in the encoding due to subjectivity. Indeed, different providers, ICUs and countries are likely to encode the information very differently. On the other hand, numerical information is in general more objective: a heart rate of 90 bpm is the same thing regardless of the clinician who recorded it or the country it was measured in. To some extent, past medical history can be summarised by scores (Charlson, Pompei, Ales, & MacKenzie, 1987; Elixhauser, Steiner, Harris, & Coffey, 1998). We used the Elixhauser score in MIMIC-III (Elixhauser et al., 1998), but diagnoses and textual past medical history were not used for state definition. We demonstrated below that textual clinical concepts and

diagnoses were to some extent encapsulated within our state definition, using the International Classification of Diseases (ICD) codes as a surrogate (see Chapter 4, section “Capture of clinical concepts and diagnoses within the states” and Figure 48).

Defining the patient cohort

Inclusion criteria

In MIMIC-III and eRI, we included all adult patients fulfilling the sepsis-3 criteria (Singer M et al., 2016). The implementation of the sepsis-3 criteria in the databases is described below. CCHIC did not include data about bacteriological sampling, so we could only rely on the administration of antibiotics and the presence of organ dysfunction to identify patients with infectious syndromes.

Exclusion criteria

- In all databases:
 - Age < 18 years old at the time of ICU admission
 - Possible withdrawal of treatment, as defined below.
 - Mortality not documented
- In MIMIC-III:
 - Intravenous fluid intake not documented
- In eRI:
 - ICU readmissions, because of the potential risk in this database of mixing up data from subsequent ICU admissions.
 - Patient admitted in an ICU with low-quality data (see below).
- In CCHIC:
 - Recorded vasopressor dose above “3”, since most corresponded to infusions in ml/h.

We excluded patients whose treatment was withdrawn, since in their case clinical decisions are no longer made aiming to optimise survival, which would have led to spurious actions in the AI policy. Withdrawal of treatment often involves patients with high severity of illness, on high doses of vasopressors, in which the treatment is withdrawn since it is considered futile. Therefore, we defined withdrawal as patients who died within 24 hours of the end of the data collection period and received vasopressors at any point and whose vasopressors were stopped at the end of the data collection. This

definition is likely to misclassify a number of patients, but after examining individual patient records, it appeared to be more reliable than using the “code status” information available in the databases.

Implementation

In MIMIC-III and eRI, sepsis was defined as a suspected infection (prescription of antibiotics and sampling of bodily fluids for microbiological culture) combined with evidence of organ dysfunction, described by a Sequential Organ Failure Assessment (SOFA) score greater than or equal to 2 (Seymour CW, Liu VX, Iwashyna TJ, & et al, 2016; Singer M et al., 2016). We adhered to the original temporal criteria for the diagnosis of sepsis: when the antibiotic was given first, the microbiological sample must have been collected within 24 hours; when the microbiological sampling occurred first, the antibiotic must have been administered within 72 hours (Seymour CW et al., 2016). The earlier event defined the onset of sepsis. In line with previous research, we assumed a baseline SOFA of zero for all patients (Raith et al., 2017; Seymour CW et al., 2016). The relevant variables were used to compute the SOFA score at each time point. The maximum SOFA score from up to 48 hours before to up to 24 hours after the presumed onset of infection was recorded and used to determine whether the patient had sepsis or not.

Since CCHIC did not include data about microbiological sampling, we could not directly implement the sepsis-3 criteria, and used antibiotics administration and a maximum SOFA score during the first 24h after admission greater than or equal to 2 to identify patients with possible sepsis. We also intended to use the Intensive Care National Audit & Research Centre (ICNARC) codes to further refine the cohort selection (proposed list of codes in Table 1), but the final anonymised dataset excluded many ICNARC codes since many of them were considered to be identifiable information.

Diagnosis	ICNARC code
Bacterial pneumonia	2.1.4.27.1
Pneumonia, no organism isolated	2.1.4.27.5
Empyema	2.1.5.27.2
Mediastinitis	2.1.5.27.3
Pleurisy	2.1.5.27.1
Mediastinitis/sternotomy related infection	2.2.3.27
Endocarditis	2.2.4.27
Peritonitis	2.3.10.27
Liver/gallbladder infection	2.3.7.27
Colon infection	2.3.6.27
Small bowel infection	2.3.5.27
CNS infection	2.4.2.27
Kidney infection	2.7.1.27
Gynaecological infection	2.7.3.27
Testes/prostate infection	2.7.5.27
Bone infection	2.10.2.27
Muscle/connective tissue infection	2.10.3.27
Skin infection	2.11.1.27
Septic shock, aetiology uncertain	2.2.12.35.2

Table 1: List of ICNARC codes of infectious diseases in CCHIC.

Exclusion of eRI hospitals with low data availability

In eRI, the data was recorded heterogeneously across ICUs. Some ICUs did not record vasopressors and/or intravenous fluids in the eRI software but used a third-party software or paper-based charts. To avoid any systematic bias in our analysis (e.g. when no medication appears in the database, where it was actually administered), we excluded hospitals for the years where the quality of data recorded was not sufficient, as data recording practices could vary over time.

We defined two separate indicators of data availability for vasopressors and intravenous fluids, averaged per day, per patient, per hospital and per year. Given that our analysis resolution is 4 hours, we expected at least 6 records to be available per day, even if the dosage was constant. Hospital-years with less than 6 daily records on average were excluded.

- Vasopressors may have been started at any time of a day, so we focused on days where we estimated that the drug was running for a whole 24-hour period. We thus selected days where the drug was running during the first 4 hours (midnight to 4 AM) and the last 4 hours of the day (8 PM to midnight). We then measured how often the administration of that drug was recorded during that same day.
- For intravenous fluids, we computed the daily average of records during the first 3 days after ICU admission, which is our period of interest, rather than averaging over the whole ICU stay. In the case of intravenous fluids, both drug delivery and no drug delivery are in general recorded.

In total, 331 ICUs out of 459 were excluded with the combined data quality selection approach (Figure 13). For comparison, the data quality was also assessed in MIMIC-III using the same definitions. MIMIC-III contained high-quality data, with a weighted average over the 5 ICUs of 20.4 intravenous fluids records and 31.1 vasopressor records per day.

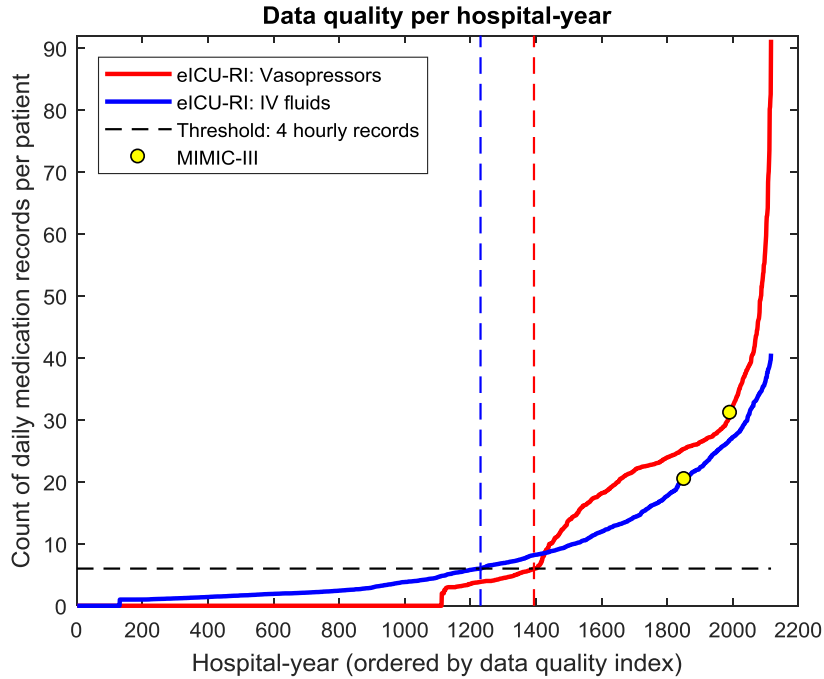
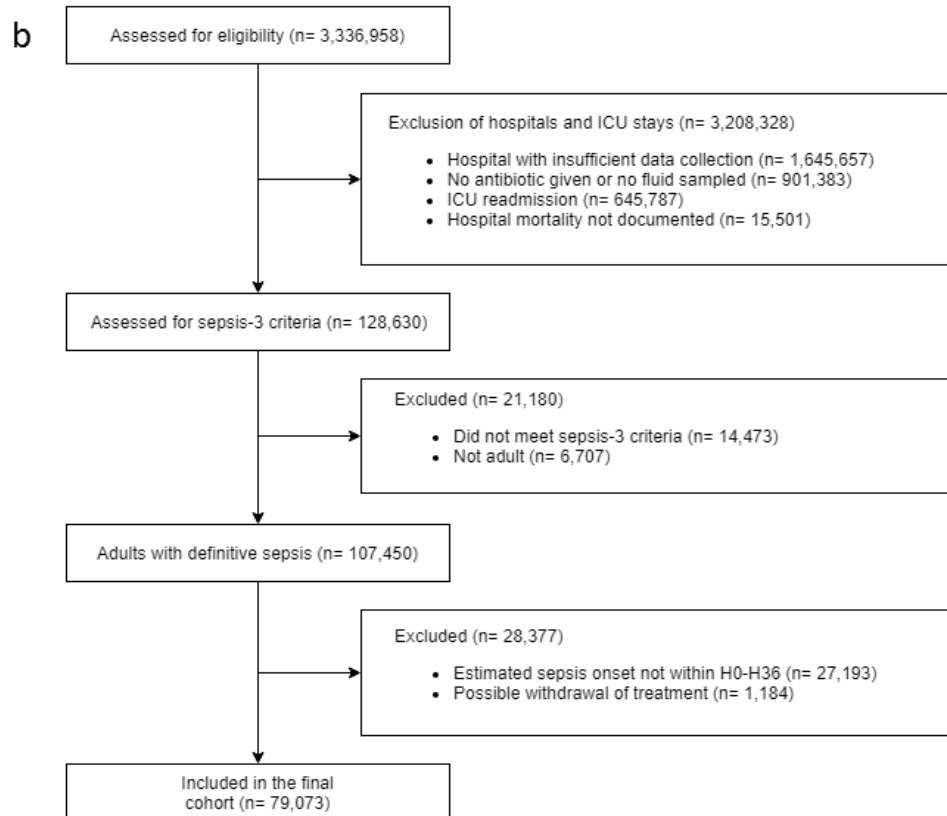
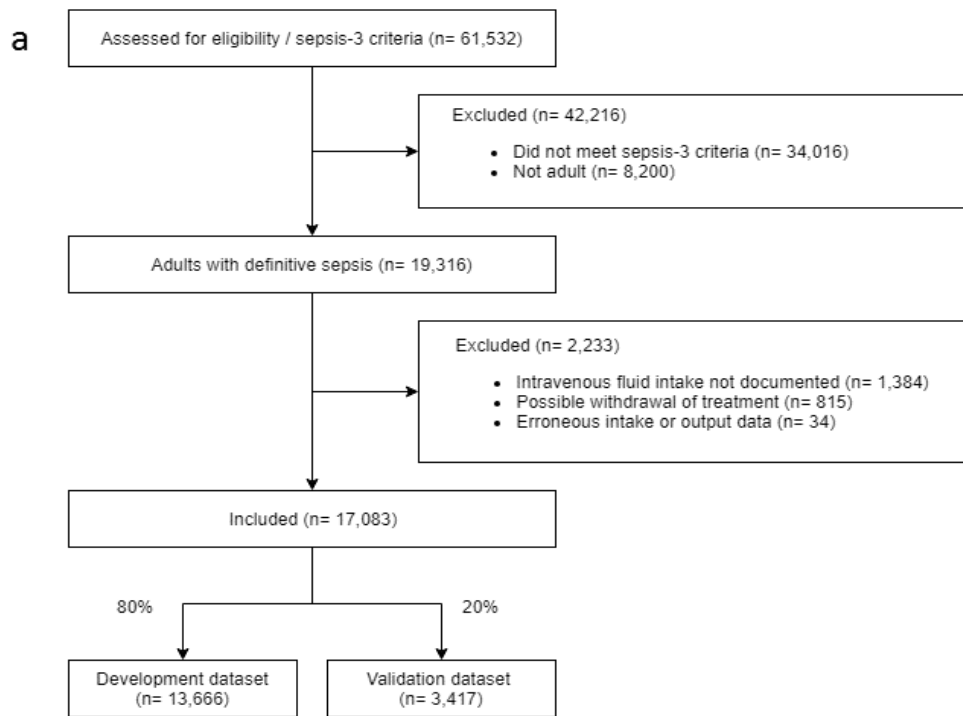


Figure 13: Estimation of data quality per hospital-year in eRI. Because some ICUs did not record intravenous fluids and vasopressors accurately, we computed the daily number of records for these 2 medications, averaged per patient, per year, in each ICU. In the figure, ICUs were ranked according to this index. We excluded all hospital-years with less than 6 daily records, to match the time resolution of the model, leading to the exclusion of 331 out of 459 ICUs. For comparison, the data quality assessed in MIMIC-III using the same definitions was also reported.

Patient inclusion flow diagrams

The patient inclusion flow diagrams are shown in Figure 14.



c

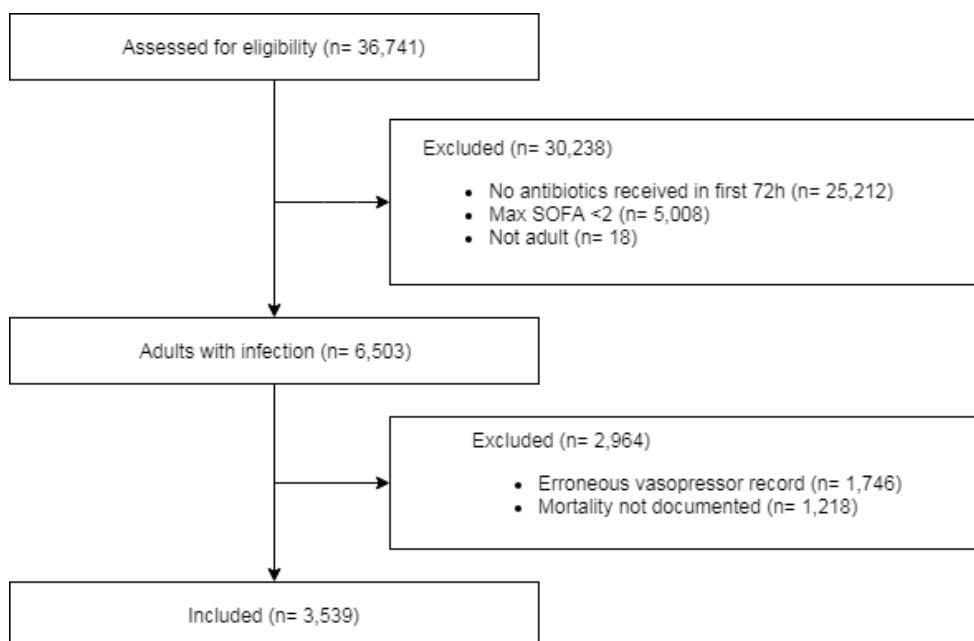


Figure 14. Patient inclusion diagrams in MIMIC-III (a), eRI (b) and CCHIC (c).

Data extraction

In all datasets, we extracted a set of up to 48 variables, including demographics, Elixhauser pre-morbid status (Elixhauser et al., 1998), vital signs, laboratory values, fluids and vasopressors received and fluid balance (see Table 2).

The first step required identifying the correct mapping of item identifiers. All three databases use a different system to encode concepts (what the data fields correspond to, e.g. heart rate, sodium, fluid balance, etc). While MIMIC-III and CCHIC use numerical item identifiers, eRI contains both structured data and free-text entries. In MIMIC-III, multiple item identifiers often map the same physiological parameter (e.g. heart rate corresponds to identifiers “211” and “220045”). Mapping the item identifiers was performed for every parameter of interest, which required knowledge of medical terminology (e.g. acronyms) including specificities of the American medical system (e.g. body temperature in degree Fahrenheit). In eRI, concepts are encoded as text fields, not as numerical item identifiers. The content of the text fields can come from drop-down lists (therefore be relatively structured) or from free-text entries (and be highly unstructured). As a result, we used regular expressions (*regex*) to retrieve information. For example, noradrenaline infusions were retrieved using the *regex* ‘%norepi%’. In CCHIC, data is encoded using proprietary NIHR Health Informatics Collaborative (NHIC) codes, whose mapping was provided by the database curators. For example, the code “NIHR_HIC_ICU_0122” corresponds to arterial lactate.

In all datasets, the data was extracted using SQL queries. The results were saved in comma-separated value (csv) files and later imported into Matlab (version 2017a) for further processing. The size of the eICU database (over 2,4 TB) precluded to employ a similar method, so the data was reformatted into 4h time steps directly in SQL. We created subqueries for each data category (vital signs, lab values, etc), which were then merged using SQL *join* queries. Extraction from CCHIC was executed in RStudio connected to an SQLite database, before the data was imported and further processed into Matlab.

Data preparation

Patients' data were coded as multidimensional discrete time series with 4-hour time steps. Data variables with multiple measurements within a 4-hour time step were averaged (e.g. heart rate) or summed (e.g. urine output) as appropriate. In MIMIC-III, fluid administration is recorded using two different formats, one being STAT doses (boluses of fluid, stored as a "total amount given"), while the other corresponds to continuous infusions (stored as a "rate" in ml/h and a "total amount given"). These inputs were converted into an amount of fluid given during each 4h time blocks. All timestamps were converted to *POSIX* time to simplify handling of date and time between the different computer programs. All features were checked for outliers and errors using frequency histograms (see example Figure 15) and univariate statistical approaches (Tukey's method). Errors were corrected when possible (e.g. conversion of body temperature from degrees Fahrenheit to Celsius). To remove further erroneous data, values above impossible thresholds were deleted (e.g. FiO2 not between 0.21 and 1, serum sodium not between 100 and 180 mEq/L, etc.).

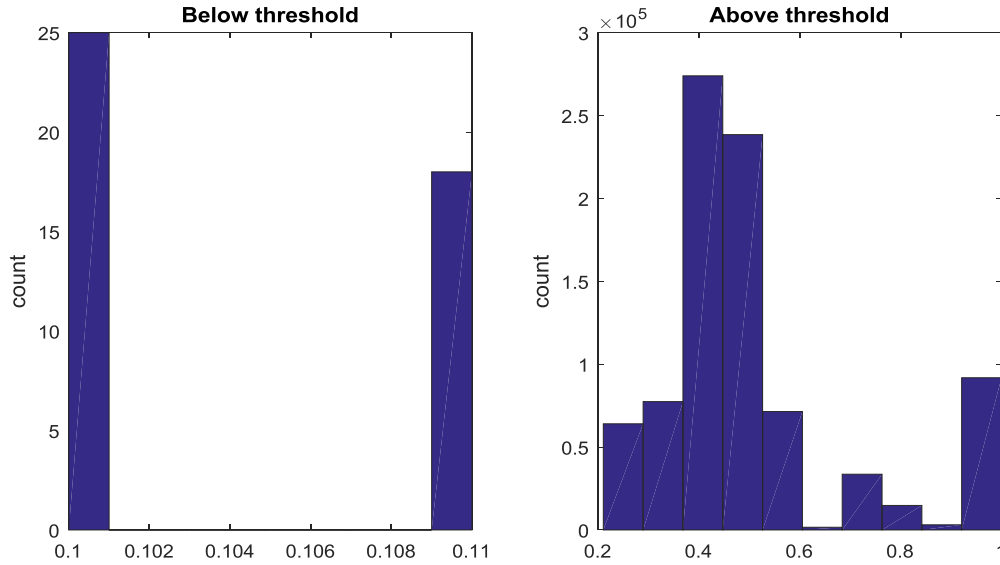


Figure 15: Illustration of the method used to identify impossible values. We show the example of the distribution of the inspired fraction of oxygen, below and above a threshold of 0.21. Forty-two values of FiO₂ of 0.1 and 0.11 correspond to impossible values (left panel) and were discarded.

Data is missing in databases for various reasons, related to patient disconnections, recording and transmission errors, or human omissions (Cismondi et al., 2013; Salgado, Azevedo, Proença, & Vieira, 2016; Tutz & Ramzan, 2015). Sometimes, the data is not missing at random, since it can be due to changes in shifts or staff-to-patient ratios, or simply because a clinician or nurse did not think that the data were important. To address the problem of missing or irregularly sampled data, we used a time-limited parameter-specific sample-and-hold approach (Hug, 2009; Kshetri, 2013). In this method, a validity period (of 2 to 24 hours) is assigned to each variable. Then, an existing value is copied in the row below if the missing value is within the validity (“hold”) period following the last available value, otherwise the data is noted as missing. In MIMIC-III, at this point, any feature with more than 50% of missing values was deleted. Also in MIMIC-III, remaining missing data were imputed using multivariable nearest neighbour imputation (Tutz & Ramzan, 2015). This was necessary since the clustering algorithm we used did not tolerate missing values. Prasad used Gaussian Process modelling for data resampling and handling of missing values, which could be a direction for future work (Prasad et al., 2017). In eRI and CCHIC, missing data were not interpolated since it was not necessary for model evaluation: only the available features were used to determine state membership of test records. The characteristics of the final datasets, with patient demographics and clinical information, are described in Table 3.

List of model features

The list of model features in the three datasets is provided in Table 2.

Category	Item	Type	Available in MIMIC-III	Available in eRI	Available in CCHIC
Demographics	Age	Cont.	+	+	+
	Gender	Binary	+	+	+
	Weight	Cont.	+	+	+
	Readmission to intensive care (binary)	Binary	+	+	-
	Elixhauser score (premorbid status)	Cont.	+	-	-
Vital signs	Modified SOFA*	Cont.	+	+	-
	SIRS	Cont.	+	+	-
	Glasgow coma scale	Cont.	+	+	+
	Heart rate, systolic, mean and diastolic blood pressure, shock index	Cont.	+	+	+
		Cont.	+	+	+
	Respiratory rate, SpO ₂	Cont.	+	+	+
Temperature					
Lab values	Potassium, sodium, chloride	Cont.	+	+	+
	Glucose, BUN, creatinine	Cont.	+	+	+
	Magnesium, calcium, ionized calcium, carbon dioxide	Cont.	+	+	-

	SGOT, SGPT, total bilirubin, albumin	Cont.	+	+	-
	Haemoglobin	Cont.	+	+	+
	White blood cells count, platelets count, PTT, PT, INR	Cont.	+	+	+
	pH, PaO ₂ , PaCO ₂ , base excess, bicarbonate, lactate, PaO ₂ /FiO ₂ ratio	Cont.	+	+	+
Ventilation parameters	Mechanical ventilation	Binary	+	+	+
	FiO ₂	Cont.	+	+	+
Medications and fluid balance	Current IV fluid intake over 4h	Cont.	+	+	-
	Maximum dose of vasopressor over 4h	Cont.	+	+	+
	Urine output over 4h	Cont.	+	+	+
	Cumulated fluid balance since admission**	Cont.	+	+	-
Outcome	Hospital mortality	Binary	+	+	+
	90-day mortality	Binary	+	-	-

Table 2: Description of the variables included in the datasets. Cont.: continuous; INR: International Normalized Ratio; * Modified SOFA: SOFA based on values in the current 4h time step; ** includes preadmission data when available; PEEP: Positive End Expiratory Pressure; PT: Prothrombin Time; PTT: Partial Thromboplastin Time; SIRS: Systemic Inflammatory Response Syndrome; Shock index: systolic blood pressure/heart rate.

Description of the datasets

Table 3 shows the characteristics of the three datasets.

	MIMIC-III	eRI	CCHIC
Unique ICUs (N)	5 (NICU excluded)	128	11
Characteristics of hospitals, per number of ICU admissions.	Teaching tertiary hospital.	Non-teaching: 37,702 (47.0%) Teaching: 29,828 (37.2%) Unknown: 12,727 (15.9%)	5 teaching tertiary hospitals
Hospital location in the USA, per number of ICU admissions.	Boston, Massachusetts.	USA: South: 32,878 (41.6%) Northeast: 15,280 (19.3%) Undocumented: 12,858 (16.3%) Midwest: 12,298 (15.6%) West: 5,758 (7.3%)	England
Type of ICUs (N, %)			
Medical-surgical ICU	-	44,567 (56.4%)	3,539 (100%)
MICU	8,131 (47.6%)	11,191 (14.2%)	-
CCU/CTICU/CSRU	4,534 (26.5%)	15,404 (19.5%)	-
SICU/TICU	4,418 (25.8%)	5,544 (7.0%)	-
Other	-	2,367 (3.0%)	-
Missing data after sample-and-hold	16.5%	9.3%	25.5%

Unique ICU admissions (N)	17,083	79,073	3,539
Unique hospital admissions (N)	17,045	79,073	3,539
Unique patients (N)	14,493	79,073	3,539
Source of hospital admission (N, %)	ED: 7,620 (44.6%) Clinic referral: 3,990 (23.3%) Transfer from external hospital: 2,760 (16.2%) Physician referral: 2,572 (15.1%) Other: 141 (0.8%)	ED: 41,241 (52.2%) Undocumented: 17,544 (22.2%) Floor: 10,753 (13.6%) Transfer from external hospital: 3,533 (4.5%) Direct admission: 2,853 (3.6%) OR: 2,988 (3.8%) Other: 161 (0.2%)	Not available
Age, years (Mean, SD)	64.4 (16.9)	65.0 (16.7)	60.5 (17.6)
Male gender (N, %)	9,604 (56.2%)	40,949 (51.8%)	58.7%
Premorbid status (N, %)			Not available
Hypertension	9,384 (54.9%)	43,365 (54.8%)	
Diabetes	4,902 (28.7%)	25,290 (32.0%)	
CHF	5,206 (30.5%)	15,023 (19.0%)	
Cancer	1,803 (10.5%)	11,807 (14.9%)	
COPD/RLD	4,248 (28.7%)	18,406 (23.3%)	
CKD	3,087(18.1%)	14,553 (18.4%)	

Primary ICD-9 diagnosis (N, %)			
Sepsis, including pneumonia	5,824 (34.1%)	41,396 (52.3%)	Not available
Cardiovascular	5,270 (30.8%)	11,221 (14.2%)	
Other respiratory conditions	1,798 (10.5%)	9,127 (11.5%)	
Neurological	1,590 (9.3%)	7,127 (9.0%)	
Renal	429 (2.5%)	1,454 (1.8%)	
Others	2,172 (12.7%)	8,747 (11.1%)	
Estimated time of onset of sepsis, after ICU admission, in hours (Median, IQR)	3.9 (-1.1 – 35.5)	1 (-0.7 – 2.8)	Not available
Initial OASIS (Mean, SD)	33.5 (8.8)	34.8 (12.4)	Not available
Initial SOFA (Mean, SD)	7.2 (3.2)	6.4 (3.5)	9.6 (3.5)
Procedures during the 72h of data collection:			
Mechanical ventilation (N, %)	9,362 (54.8%)	39,115 (49.5%)	3,322 (93.9%)
Vasopressors (N, %)	6,023 (35.3%)	23,877 (30.2%)	3,322 (93.9%)
Renal replacement therapy (N, %)	1,488 (8.7%)	6,071 (7.7%)	Not available

Fluid balance on admission documented (N, %)	9,317 (54.5%)	24,672 (31.2%)	0%
Length of stay, days (Median, IQR)	3.1 (1.8 – 7)	2.9 (1.7 – 5.6)	2.5 (1.1 – 5.2)
ICU mortality	7.4%	9.8%	Not available
Hospital mortality	8.9%	16.4%	15.3%
28-day mortality	11.3%	Not available	Not available
90-day mortality	18.9%	Not available	Not available

Table 3: Description of the datasets. CCU: Coronary Care Unit; CHF: Congestive Heart Failure; CKD: Chronic Kidney Disease; COPD: Chronic Obstructive Pulmonary Disease; CSRU: Cardiac Surgery Recovery Unit; CTICU: Cardio-thoracic ICU; ED: Emergency Department; ICD-9: International Classification of Diseases version 9; IQR: Interquartile Range; MICU: Medical ICU; NICU: Neonatal ICU; OASIS: Oxford Acute Severity of Illness Score; OR: Operating Room, RLD: Restrictive Lung Disease; SD: Standard Deviation; SICU: Surgical ICU; SOFA: Sequential Organ Failure Assessment; TICU: Trauma ICU.

Discussion on dataset preparation and patient cohort definition

Several aspects of the way the cohorts were defined and the datasets prepared can be discussed.

The chosen time resolution of 4 hours represents a trade-off between the ability to represent acute clinical changes (a shorter time resolution is better) and dealing with data with low sampling frequency. If the time resolution was very short (e.g. one minute), then consecutive rows would mostly often be identical since we used the sample-and-hold method for missing values interpolation. If the time resolution was much longer (e.g. 24 hours per time step), then the model would not be suitable to capture acute changes in patient physiology. It is likely than a time resolution of 1 or 2 hours would equally have worked.

In line with previous research, we assumed a baseline SOFA of zero for all patients (Raith et al., 2017; Seymour CW et al., 2016). In MIMIC-III, we also tested the impact of taking into account past medical history in the SOFA score using the following method: we estimated whether a patient had a history of

COPD/emphysema, CKD, liver disease or thrombocytopenia (from the ICD codes) and discounted the SOFA points related to these conditions. When removing the SOFA points in patients with past medical history, 135 patients only (out of 17,898) went from sepsis positive to sepsis negative, representing less than 1% of the cohort. For simplicity, this was not implemented when defining the final cohort.

We computed the SOFA score for each 4 hour time interval, which strictly speaking differs from the original SOFA definition, where the score is computed using the worst values over the last 24 hours (J. L. Vincent et al., 1996). In MIMIC-III, all but 2 patients with suspected infection reached a maximum SOFA of 2 or more points. Therefore, the equation $sepsis = suspected\ infection\ in\ the\ ICU + SOFA \geq 2$ could be simplified to $sepsis = suspected\ infection\ in\ the\ ICU$ in our cohort. We followed the international recognised sepsis definition in which timing of sepsis onset is not precise since it depends on what time the interventions (administration of antibiotics or sampling of body fluids) were recorded, which obviously differs from the actual onset of sepsis at a biological and cellular level, which is impossible to detect. Left censoring was present or suspected in many patients whose onset of sepsis and initial resuscitation occurred prior to ICU admission and data collection. We only included patients in which fluid samples and antibiotic administration was recorded. This was a deliberate conservative approach to define the cohort to ensure all patients had sepsis. It is likely that many patients who actually had sepsis but whose data was missing were not picked up with this method.

We were unable to directly implement the sepsis-3 criteria in the CCHIC database, so the cohort definition differs across the datasets, which is a limitation of the work. We argue that sepsis definitions remain highly inconsistent across definitions and implementations. For example, Shankar-Hari identified a 3.5-fold drop in septic shock in a single cohort with sepsis-3 criteria compared with sepsis-2 (Shankar-Hari et al., 2017).

Chapter summary

To summarise, this chapter described how we defined our cohort, the process used for data extraction and the preparation of the datasets. The following chapter will describe how the MIMIC-III training dataset was used to define all the elements of the MDP.

Chapter 3: Model construction

This chapter describes how the MIMIC-III training dataset was used to build an MDP on which the RL algorithms were deployed.

Choice of model

The complexity of the ICU environment and sepsis and its treatment, with time-varying exposure and outcome, and the impossibility to isolate a single instrumental variable render difficult alternative techniques such as instrumental variable analysis or simpler propensity scoring (D’Agostino RB, 2007).

Causal inference methods (such as targeted maximum likelihood estimate or marginal structural models) and Markov models are two different approaches to evaluate the effect of decision strategies, both of which have their strengths and limitations (Gruber & Laan, 2009; Maldonado & Greenland, 2002). Reconciling causal inference and reinforcement learning is an area of ongoing research.(Gershman, 2017) While causal inference focuses on the single decision taken having observable or potentially unobserved causal effects, reinforcement learning focuses on the overall decision strategy, which is key to the AI clinician innovation - it captures the cognitive capabilities of clinicians to pick a treatment action, observe the patient response, adjust or continue its treatment accordingly, etc. For these reasons, we selected Markov models for this research.

The true patient physiological state is only partially observable using physical and biological measurements. This observability is further restricted by the subset of measures actually taken in the ICU, which limits both the nature and frequency of data recorded. Strictly speaking, and as discussed in the first chapter of this thesis (section “Introduction to reinforcement learning and notation”), modelling the decision-making process should be mathematically formulated as a Partially Observable Markov Decision Process (POMDP), where the treatments given to patients correspond to actions (Bennett & Hauser, 2013; Puterman, 1994; Spaan, 2012; Sutton & Barto, 2018). Computing the optimal POMDP solution from the available data would require performing inferences of probability distributions over potential true patient states as well as account for all possible continuous treatment actions. This would be computationally impractical given the size of the state and action spaces (although new methods to achieve this are being explored (Li, Komorowski, & Faisal, 2018)), so we applied a two-step simplification when building our model.

First, we assumed that the state of the patient consisted only of the information that was recorded in the ICU electronic medical record. Thus, the POMDP reduced to a Markov Decision Process (MDP), which

we refer to as the “full MDP”. At this stage, the MDP contains continuous states and actions: the number of possible states and actions is infinite, which makes it difficult to solve. Solving a continuous MDP can be done in two ways: discretisation and value function approximation (Undurti, Geramifard, & How, 2011). Discretizing a continuous MDP can be done by “gridding” the state and action spaces. Quantizing state and action spaces then apply finite-state methods is a common method to obtain approximate solutions of tasks with continuous states and actions (Sutton & Barto, 2018). For example (Figure 16), a two-dimensional continuous space can be discretized using a grid:

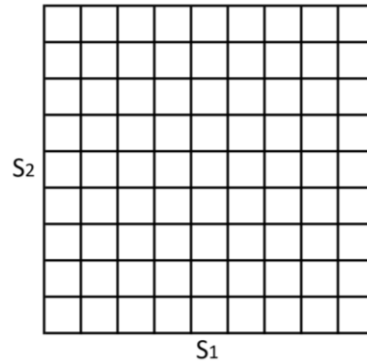


Figure 16: Discretising a two-dimensional (S_1, S_2) state space into a 9x9 grid.

The “curse of dimensionality” precludes using this approach on our dataset. Suppose we discretise each of the n dimensions of the state into k values, the total number of discrete states we obtain is k^n . For example, if we choose to discretise each of the 45 continuous variables of the model (see Table 2) into 100 bins, we would have $100^{45} = 10^{90}$ possible states, which is more than the estimated number of atoms in the known universe, far too many to be solvable.

Another possible approach to find policies in continuous MDP (which we did not implement) is to approximate the value function directly without resorting to discretisation or state aggregation, using the so-called value function approximation. This is commonly achieved using the fitted value iteration algorithm, where we approximate V as a function of the states, using any regression algorithm (linear or non-linear):

$$V(s) = \theta^T \phi(s)$$

With ϕ is some appropriate feature mapping of the states.

A more practical approach (which we applied) is to aggregate the data points by their similarity, a process called clustering. The intuition behind this state aggregation is that we consider similar

measurements to be the same, for example that the difference between a blood pressure of 120/80 and a blood pressure of 120/78 is not clinically relevant.

Second, we discretised the doses of vasopressors and intravenous fluids into finite actions, each corresponding to a dose range of the two medications. Quantizing the action space is also a common approach (Sutton & Barto, 2018).

Defining the elements of the Markov decision process

We defined a Markov decision process (MDP) as a tuple $\{S, A, T, r, \gamma\}$ where S is a set of states, A a set of actions, T a transition matrix, r a reward function and γ a discount factor. In this section, we describe how those various elements were defined, starting from the raw clean MIMIC-III training data.

Defining states

We intend to model and analyse the dynamics of patients as they evolve among various clinical states. The discrete state space was defined by clustering all patient time series from the MIMIC-III development set.

Various clustering algorithms could allow us to group matching patients together. A good cluster hierarchy is one in which individuals that are in the same cluster are similar with respect to their observable properties. We combined all patient time series from the MIMIC-III development set into a single array, which was clustered using k-means. Clustering algorithms such as k-means are a class of unsupervised machine learning tools (Bishop, 2007). The algorithm assumes that the data lives in an N -dimensional Euclidean space, and requires a parameter k , the desired number of clusters.

We start with randomly initialised cluster centres C_k . The algorithm alternates between 2 steps: 1) the assignment (or expectation) step, where each data point is assigned to the closest cluster C_k ; 2) the refitting (or maximisation) step, where each cluster centre is moved to the centre of gravity of the data assigned to it (Figure 17):

$$C_k \leftarrow \frac{1}{N_k} \sum_{i \in C_k} X_i$$

With N_k the data dimensionality and X_i the data points.

By following these 2 steps until convergence (when the cluster assignment stabilises), the algorithm generates k clusters minimising the intra-cluster variance and maximising the inter-cluster variance.

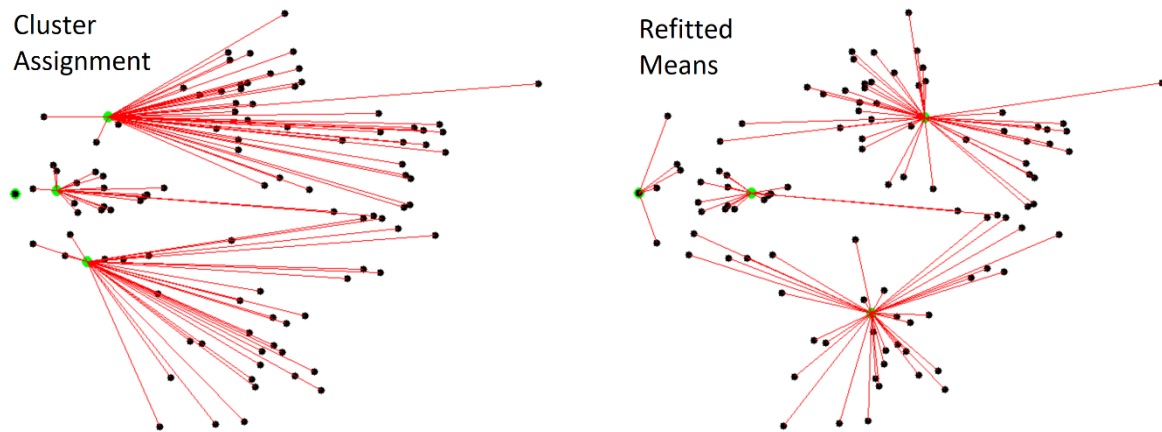


Figure 17: Conceptual illustration of the 2 steps of the k-means algorithm. After random initialisation of k cluster centres C_k , the algorithm alternates between 2 steps: 1) the assignment step, where each data point is assigned to the closest cluster C_k ; 2) the refitting step, where each cluster center is moved to the centre of gravity of the data assigned to it. After convergence, this produces k clusters. Illustration credit: Andrey A. Shabalin.

K-means was used because it is computationally efficient (it produces a candidate clustering solution of the whole training set in around 5 seconds) and because – by definition – it produces discrete groups of matching, homogeneous patients. We used k-means++, an improved version of k-means where the initial values of cluster centroids (or "seeds") are not chosen totally at random (Arthur & Vassilvitskii, 2007). To homogenize the weight of variables before clustering, the variables with a normal distribution were standardised. Log-normal distributed variables were log-transformed before standardisation. The binary data were centred, so their mean was zero. The normality of each variable was tested with visual methods: quantile-quantile plots and frequency histograms. The dataset does not contain categorical non-binary data.

More advanced clustering algorithms could have been employed, some of which represent promising alternatives for future model development, for example non-linear time series clustering and *kamila* (Foss & Markatou, 2018; Zhang & An, 2018). Non-linear time series clustering include for example autoregressive and moving average models, and bypass the assumption that the time series are only

linearly dependent (which usually fails in practice) (Zhang & An, 2018). *kamila* allows clustering of mixed-type data and combines equitably continuous and categorical variables without requiring strong parametric assumptions (Foss & Markatou, 2018).

It is likely that our model would work for a range of values of k , the number of clusters. If we choose a very low number (say, a dozen of states), then the model granularity would be too coarse and unable to capture mild changes in patient physiology. With a very high number of states (say, many thousands), then the average state population would be too low and the transition matrix would become very sparse. The chosen number of clusters k was 750 as that number minimised both the Akaike and Bayesian Information Criterion (Figure 18). Information criteria are likelihood criterion penalised by the model complexity (the number of parameters in the model), and represent the intrinsic complexity present in the data (Akogul, Erisoglu, Akogul, & Erisoglu, 2017; Hu & Xu, 2003). The literature recommends selecting the clustering solution minimising information criteria, since they represent the clustering solution that maximises the likelihood function of the model (Akogul et al., 2017; Hu & Xu, 2003; Jones, 2011).

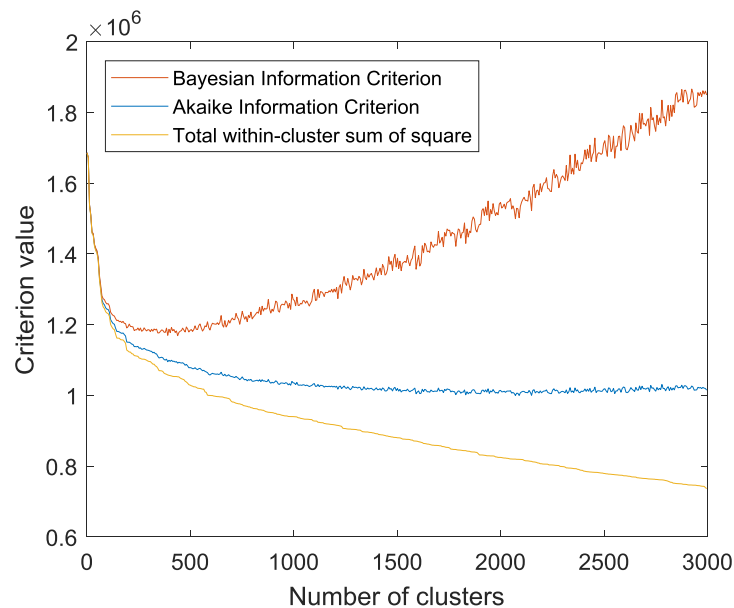


Figure 18: Selection of the number of clusters in the model. Bayesian information criterion (BIC), Akaike information criterion (AIC), and total within-cluster sum of square were computed for a range of k , the number of model clusters (states). The minimum is found for k around 2,000 for AIC and $k=400$ to 500 for BIC.

In all the 500 models that we built (please refer to Figure 12 for further explanations), the clustering solution that maximised the inter-cluster variance and minimised the intra-cluster variance was chosen among 32 candidate clustering solutions. Figure 19 shows a low-dimensional projection of the cluster centroids (the “health states”) on the first 3 principal components of the data, for the final model. The median number of state visits (how many times a particular state is visited by any patient) is 226 (IQR 112-356), as shown in Figure 20. Two additional terminal absorbing states were added to the state space, corresponding to discharge (success) and death (failure), and added at the end of each sequence depending on the actual patient’s outcome.

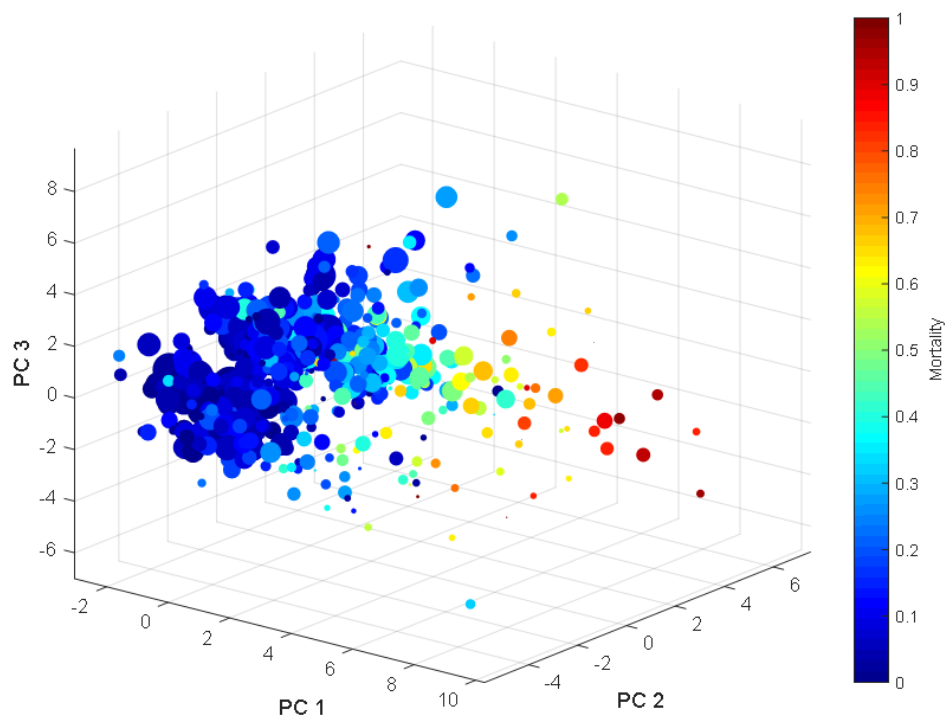


Figure 19: Visualisation of the cluster centroids projected on the first 3 principal components of the data. Each dot corresponds to a cluster, whose size and colour correspond to, respectively, the number of patients in the state and the average state mortality.

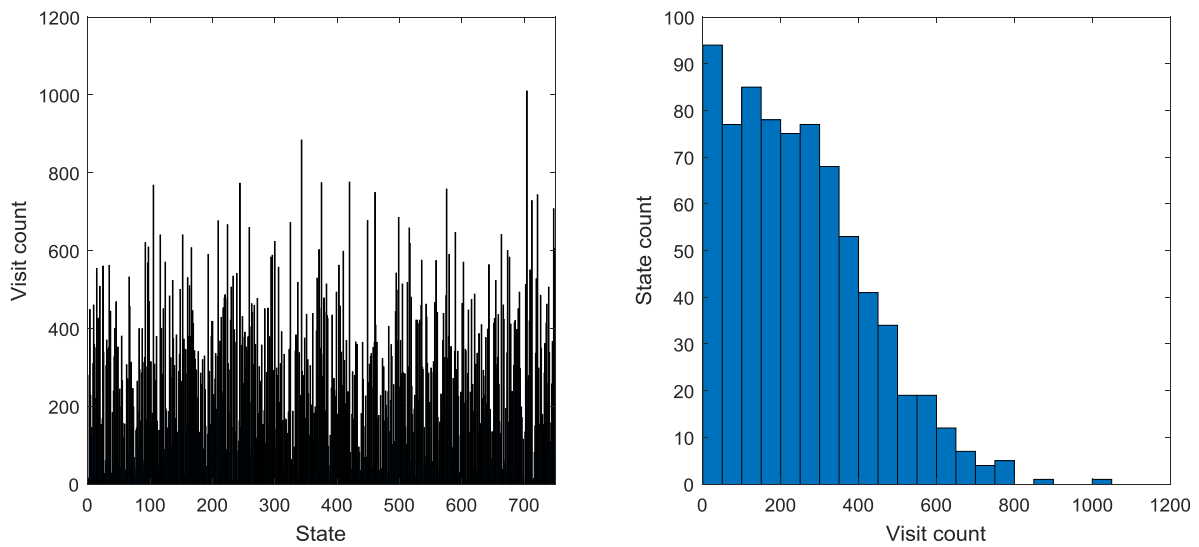


Figure 20: State visit counts (left) and distribution of the state visit counts (right) in the MIMIC-III training data. The median number of state visits is 226 (IQR 112-356).

We tested the robustness of our state definition method to missing data. In the eRI cohort, we added increasing proportions of missingness at random (0 to 20%), and observed how the state membership would be affected (Figure 21). The missing data was added to the whole eRI dataset, by removing random data points to any column at random (see formula below, with missingness $\epsilon \in [0 - 0.2]$; *NaN* stands for “not a number”).

```
dataset(rand(size(dataset))<missingness)=NaN
```

With 10% of missing data, 28% of the records ended up in a different state, and 43% with 20% of missing data. Then, we assessed which variable would affect the most the state membership. We added 50% of missing data at random to each of the 46 variables of the eRI dataset and measured how state membership was affected. Figure 22 shows the variables ranked according to their effect on the state membership. Important features (at the top of the list) lead to a larger drop in the correct state membership assignment. This is important information for two reasons: 1) extra care must be placed on the data quality of the predominant features; 2) for feature selection (if we wanted to develop a sparse model), it would be important to keep the most predominant parameters.

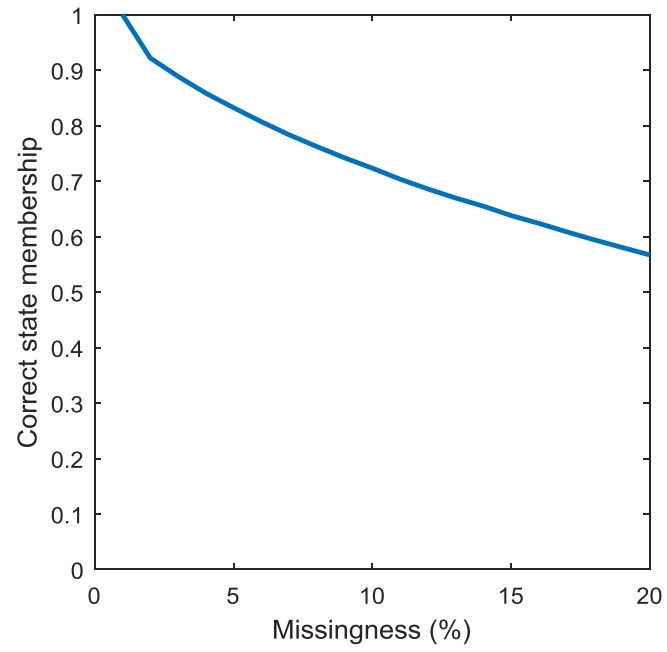


Figure 21: Robustness of the state definition to missing data, in the eRI cohort. We observed how many records would remain in the correct state after adding increasing proportions of missing data at random.

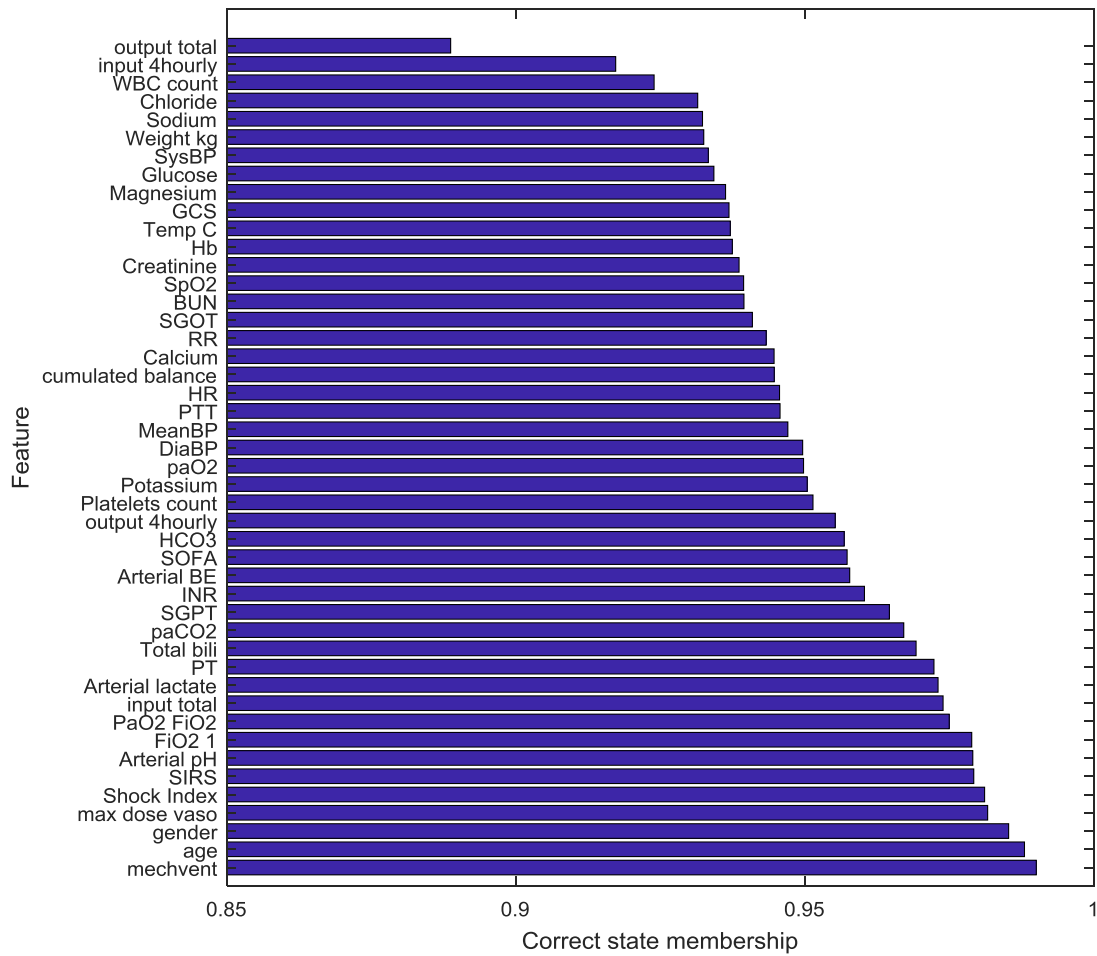


Figure 22: Importance of the model features on the state membership. We added 50% of missing data at random to each of the 46 continuous variables of the eRI dataset and measured how many records would remain in the correct state. In this analysis, important features (at the top of the list) lead to a larger drop in the correct state membership.

We related this analysis to the actual proportion of missing data in eRI, which is 9.3% globally. However, the proportion of missingness in the most important features is reassuringly low (Table 4), which means that the state membership attribution method is most likely robust for eRI.

Feature	Importance ranking	Proportion of missingness (%)
Output total	1	0
Input 4 hourly	2	0
WBC count	3	0.92
Chloride	4	0.82
Sodium	5	0.78
Weight	6	1.39
Systolic BP	7	2.57
Glucose	8	0.87
Magnesium	9	16.96
GCS	10	4.09

Table 4: Proportion of missingness in the eRI cohort for the most important features for state membership attribution, as identified by the method above. Reassuringly, missingness is low for the top features.

Other dimensionality reduction techniques can be tested to help visualising the dataset. t-Distributed Stochastic Neighbor Embedding (t-SNE) is a relatively new technique for dimensionality reduction, that is particularly well suited for visualisation of high-dimensional datasets (Maaten & Hinton, 2008). In a nutshell, t-SNE models each high-dimensional object by a two-dimensional point in such a way that similar objects are modelled by nearby points and dissimilar objects are modelled by distant points with high probability (Maaten & Hinton, 2008). t-SNE is very different from k-means in the sense that it does not preserve distances nor density, and only to some extent preserves nearest-neighbours. It requires fine tuning of a hyperparameter called “perplexity”, which loosely speaking can be viewed as a knob that sets the number of effective nearest neighbours. A larger or denser dataset is expected to require a larger perplexity. Figure 23 shows an example of 10,000 data points from the development dataset processed with t-SNE, for two different values of the perplexity. Data points are colour-coded according to their k-means state membership. In the figure, we see that t-SNE often groups together patients belonging to the same cluster according to k-means, which corroborates the k-means approach. Indeed, if two different clustering methods put data points together, it reinforces our confidence that they do resemble each other.

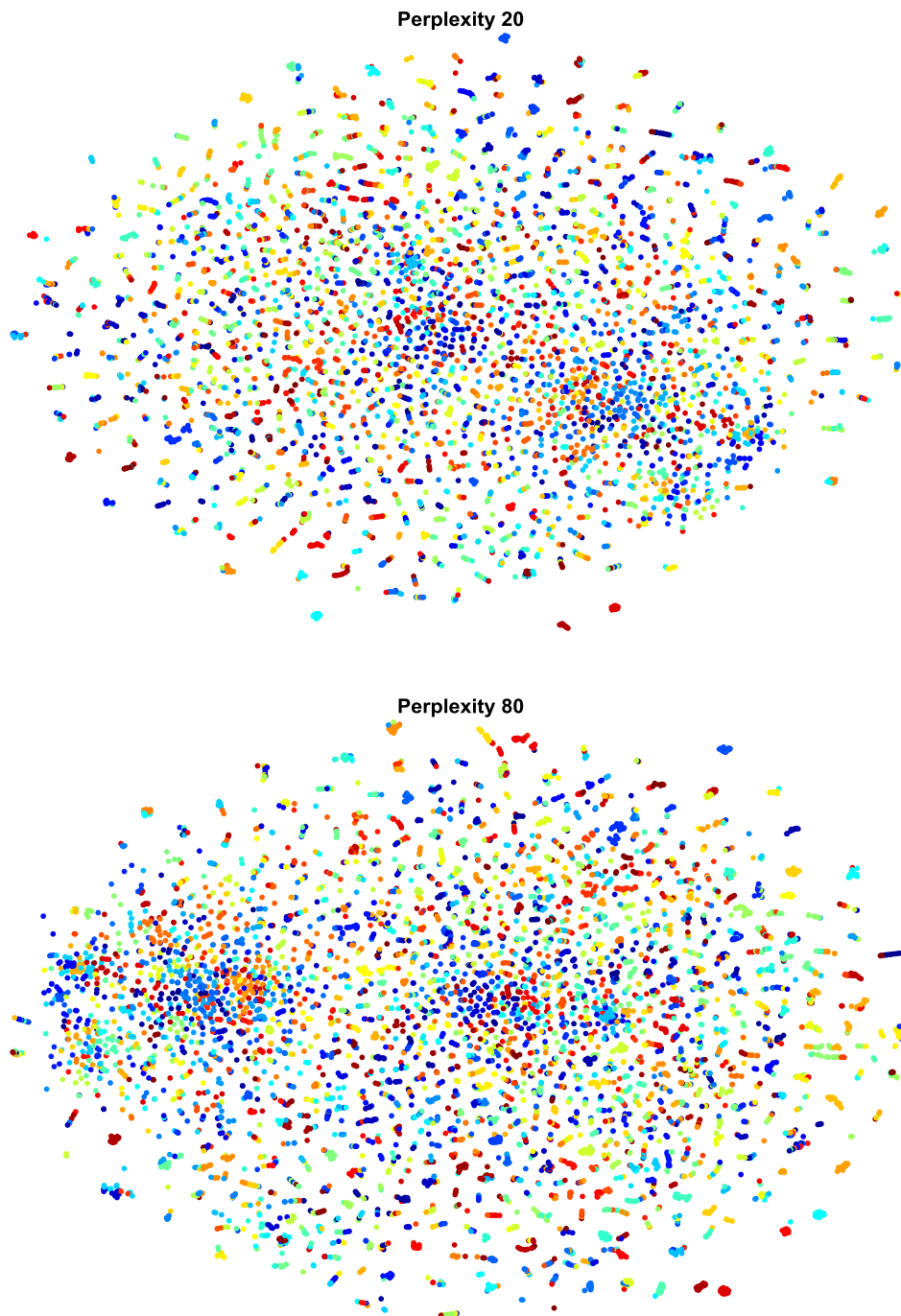


Figure 23: Visualisation of 10,000 patient data points using t-SNE, for 2 different values of the perplexity hyperparameter (top: perplexity 20, bottom perplexity 80). Loosely speaking, perplexity can be viewed as a knob that sets the number of effective nearest neighbours. Colours represent clusters as defined by k-means++. Reassuringly, t-SNE often groups together patients belonging to the same cluster according to k-means (especially easy to visualise at the periphery), which validates the k-means approach.

Defining actions

We defined the action space from the drugs administered to patients in the MIMIC-III training sample. The management of ICU patients with sepsis is extremely complex and includes several core principles such as rapid control of the source of infection including treatment with antibiotics, correction of hypovolaemia, and management of secondary organ failures including mechanical ventilation, renal replacement therapy, sedation and nutritional support to name just a few. Including all these potential interventions as actions in the MDP would have required a much larger dataset. A key challenge is arguably the management of intravenous fluids and vasopressors, as discussed in the introductory chapter. Consequently, we focused on medical decisions regarding the total volume of intravenous fluids and maximum dose of vasopressors administered over each 4h period.

Intravenous fluids included boluses and background infusions of crystalloids, colloids and blood products, normalized by tonicity as previously described (Waechter et al., 2014). The vasopressors included norepinephrine, epinephrine, vasopressin, dopamine and phenylephrine, and were converted when necessary to norepinephrine-equivalent using previously published dose correspondence (Brown et al., 2013). In MIMIC-III, the dose of vasopressors and intravenous fluids were zero for 83% and 23% of the samples, respectively.

Figure 24 shows the distribution of the non-null drug doses. To define the action space, the dose of each treatment was represented as one of five possible choices, choice 1 being “no drug given”, and the remaining non-null doses divided into four quartiles (Table 5). The permutation of the two treatments produced 25 possible discrete actions (Figure 25). We expressed the suggested dose as the median of each dose bin matching a suggested action. CCHIC does not include IV fluids data. For this reason, the evaluation was limited to direct methods (please refer to Chapter 4, section “direct policy assessment methods”), since we could not fully determine which actions had been taken by clinicians.

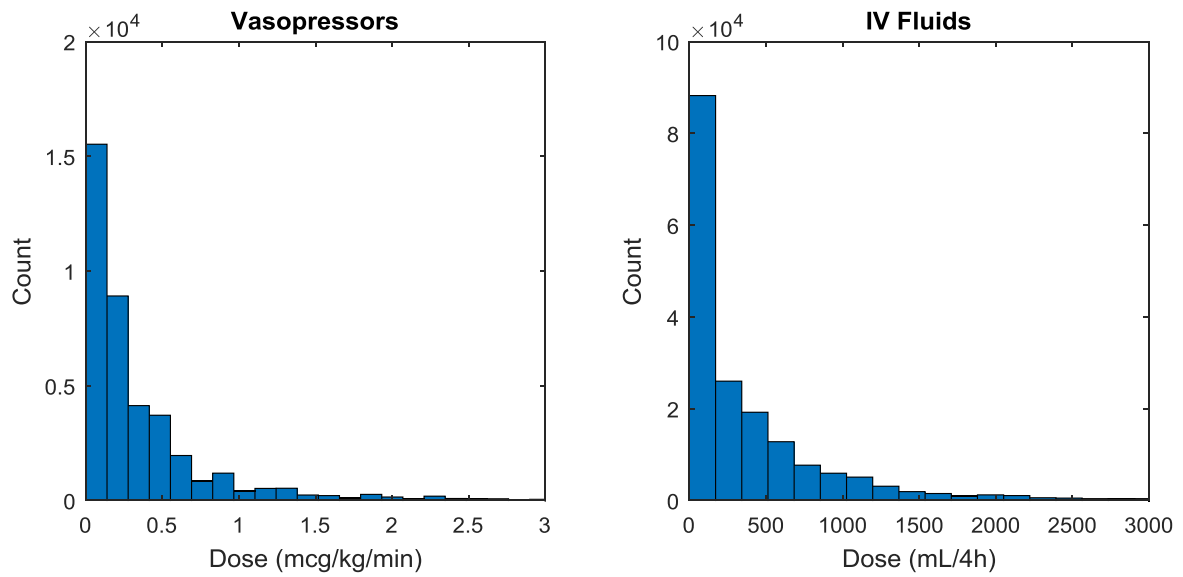


Figure 24: Distribution of non-null doses of vasopressors and intravenous fluids in the MIMIC-III dataset.

Discretized action	IV fluids (mL in 4 hours)			Vasopressors (mcg/kg/min)		
	Range	Median dose	Proportion (%)	Range	Median dose	Proportion (%)
1	0	0	23.30	0	0	83.07
2]0-50]	30	18.52]0-0.08]	0.04	4.07
3]50-180]	85	19.82]0.08-0.22]	0.13	4.39
4]180-530]	320	19.19]0.22-0.45]	0.27	3.94
5	>530	946	19.15	>0.45	0.68	4.52

Table 5: Range and median doses of drugs for the discretized actions.

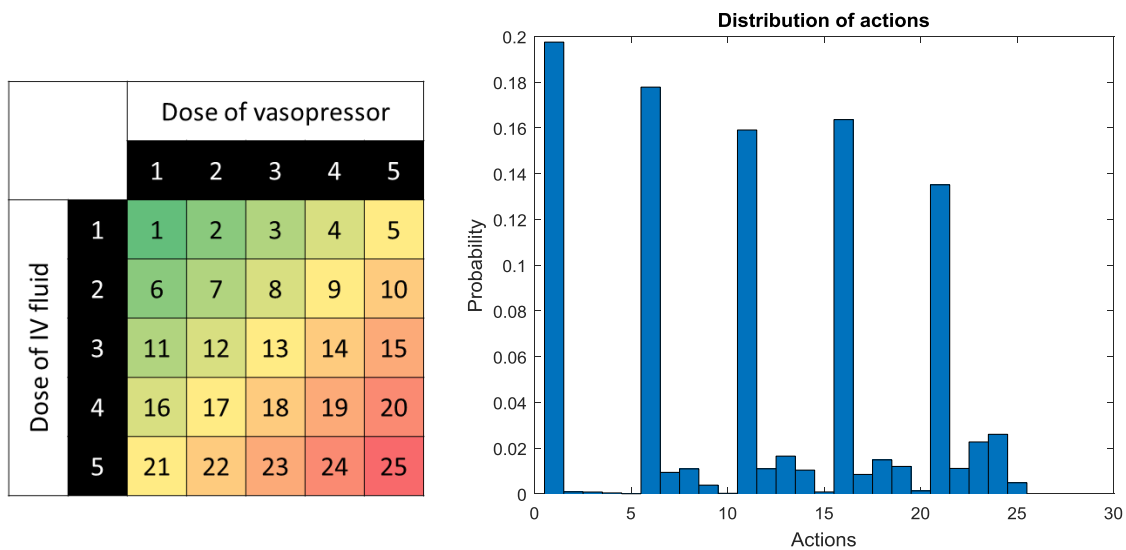


Figure 25: Representation of the action space (left) and distribution of actual clinicians’ actions (right). Note how actions with zero vasopressor (1, 6, 11, 16 and 21) are over-represented.

Transition matrix & Markov property

We used the framework of a stochastic MDP, where the next state was not deterministically given by the current state and action. This matched the clinical setting, where the same treatment applied to a group of similar patients has a range of effects and leads the patients to various subsequent health states. Formally, the transition matrix $T(s', s, a)$ contained, in each cell, the probability to end up in s' when, starting from state s , the action a was taken. It forms a k -by- k -by- c matrix, with k states and c actions. It was computed by counting of how many times each transition was observed in the MIMIC-III training samples. Then, the 3D matrix containing the counts was divided by $\sum_{i=1}^N T(:, S_i, :)$ with N = number of states, which is the sum of all possible transitions from state s . Existing literature used a similar approach (Welton, Sutton, Cooper, Abrams, & Ades, 2012). Figure 26 shows the distribution of the occurrences of all the transitions observed in the MIMIC-III training data, with a state space of 752 and an action space of 25. 72.5% of the transitions (13,626 out of 18,800) are never observed. Nearly a third of the observed transitions (31.8 %) are seen only once, and close to two thirds (65.8 %) are seen 5 times or less.

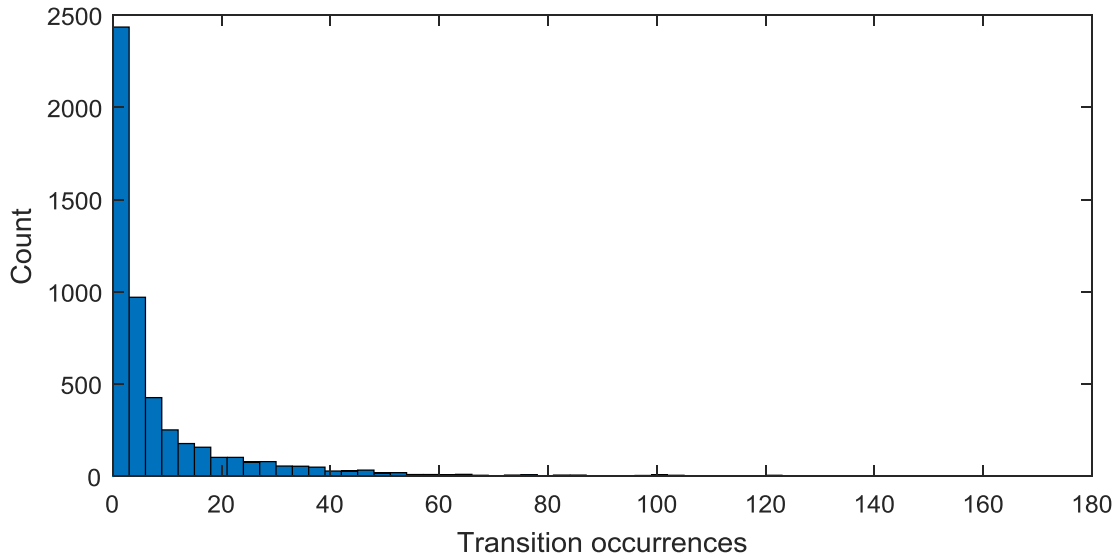


Figure 26: Distribution of the occurrences of all the transitions observed in the MIMIC-III training data, with a state space of 752 and an action space of 25. 72.5% of the transitions (13,626 out of 18,800) are never observed, while another 8.8% (1,643 out of 18,800) are seen only once.

It is very likely that some of the rarest transitions represent noise in the data (e.g. impossible transitions) or real but improbable events (e.g. sudden and unexpected improvement or deterioration in health). The presence of such transitions is likely to corrupt the AI agent, since it will always choose decisions likely to lead to favourable transitions to better health states, even if the probability of such a transition happening is extremely low. The AI agent may recommend actions associated with a rare transition that was observed in the data, but in reality highly unlikely to genuinely exist. We have described this issue as the “dirty gridworld” problem, in reference to the artificial gridworld tasks used in the machine learning literature (Figure 27). We would argue that the sepsis (as we formulated it) is indeed a (dirty) gridworld problem. When analysing individual patient trajectories (see Chapter 5, “model testing on eRI”), we demonstrate that an additional problem is represented by right censoring of patient data, where a patient is severely ill at the end of the data collection period (48h after the estimated sepsis onset) but survives at 90 days. The final action will be rewarded highly, when in reality it had little effect on the end outcome and should have been discounted. The mirrored phenomenon is expected to exist for apparently healthy patients at the end of the data collection period but who did not survive to 90 days.

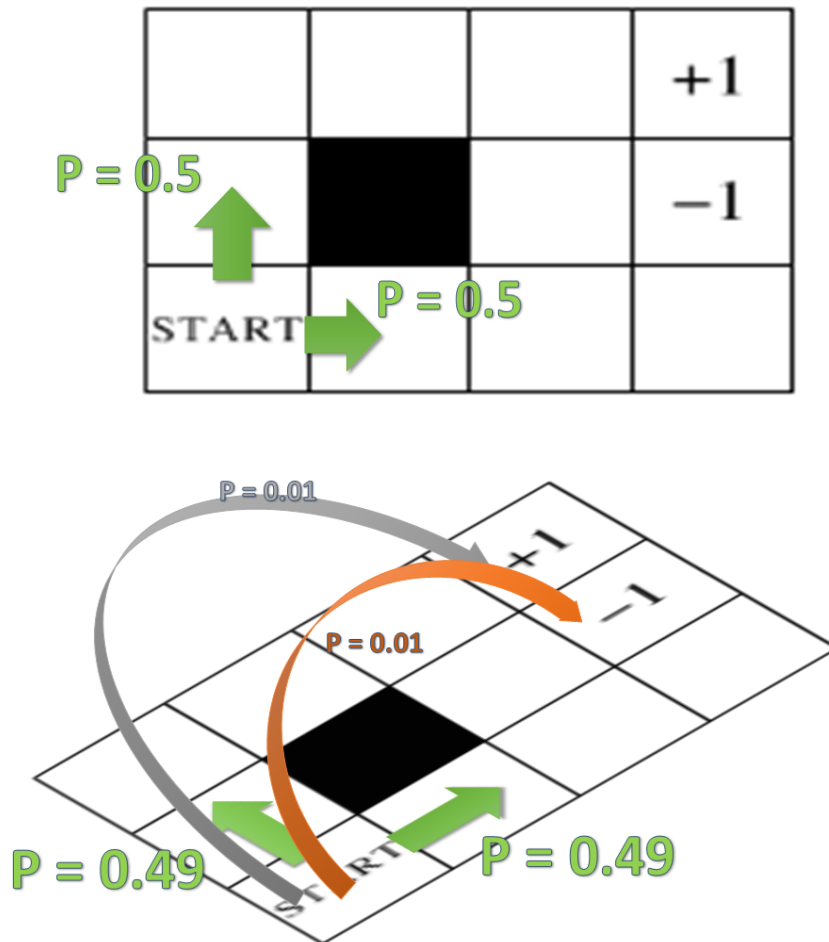


Figure 27: The dirty gridworld problem, which illustrates how a noisy transition matrix may corrupt the behaviour of the AI agent. In a “clean” gridworld (top), transitions are only possible into adjacent states. In this example, the only two possible transitions from the starting state are going up or right, each with a 50% probability. In a “dirty” gridworld (bottom), it may be possible to reach most states (including the desirable state of being discharged alive) from any other state. If discharge can be reached from any state, the actions leading to these transitions become recommended even if highly unlikely or erroneous (if they are impossible in real life).

A partial solution to address the dirty gridworld problem is to “prune” the transition matrix and eliminate the transitions seen rarely. We discarded all the transitions that were seen 5 times or less in the data (they represent 1410 out of the 5174 existing transitions, a 27% reduction). This increased the sparsity of the transition matrix (the fraction of transitions that are never observed) from 72.5 % up to 80 %. As such, the AI agent will now only recommend actions that were taken by actual doctors a fair number of times. Also, by pruning the transition matrix, some rare transitions into remote states disappeared, so that these clusters became isolated from the bulk of the state space, with no possibility

to transition in or out of them. An evaluation of the transition matrix revealed that this was the case for 26 states, but that they were very sparsely populated states which contained altogether only 388 data points (the median state population from the training set is 226 data points).

By pruning the transition matrix, we turn the AI agent from a free-roaming agent that can just recommend any action into an agent that selects the most optimal action among actions taken frequently by clinicians. This manipulation is likely to have increased the safety (hence the acceptability by patients, clinicians and other stakeholders) of the AI model. To visualise the effect of this manipulation, we show in Chapter 5 the distribution of actions suggested by the AI agent with and without pruning the transition matrix (Figure 56 and Figure 57).

MDP models rely on the Markov property of the states, i.e. the chain is memoryless and all the information is encapsulated in the current state. This property can be mathematically verified if the time to remain in a given state is exponentially distributed (Norris, 1997). In order to verify this, the lifetime in each state was computed by bootstrapping, then an exponential decay function $y = a \cdot e^{b \cdot x}$ was fitted to the distribution of the life time (Figure 28, left). The correlation between the life expectancy and the fitted exponential function was computed in each state. The distribution of the correlation in all states was used as our primary indicator that the model satisfies the Markov property (Figure 28, right). The high median correlation coefficient r^2 of 0.97 (IQR = 0.1) confirms that the life expectancy in most states follows an exponential decay: our chain is Markov. Each subsequent state, action and reward depends on the previous step only, but not on the previous full history.

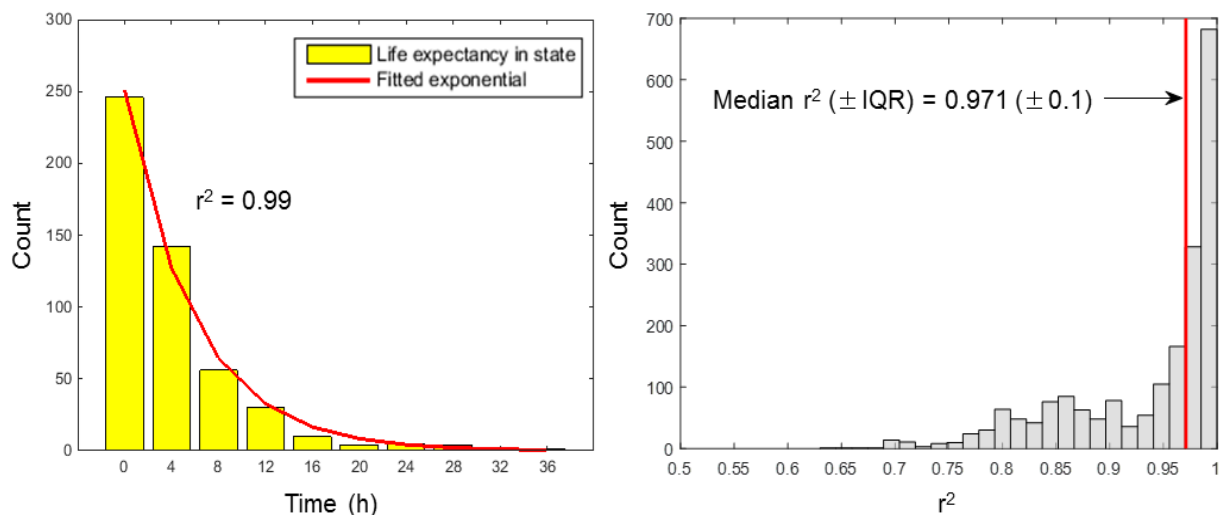


Figure 28: Verification of the Markov property. We measured the life expectancy in each state, by observing how long an agent would remain in a given state when following the transition matrix, in 500 trials. a, Example of the life expectancy in one state, with fitted exponential decay function. The correlation coefficient r^2 between the data and the fitted function is 0.99 in this example. b, Distribution of the correlation coefficients between life expectancy and exponential decay functions in the 750 states of the model. The high median correlation coefficient of 0.97 (interquartile range 0.1) confirmed that the life expectancy in most states was indeed memoryless.

Even if the chain is mathematically memoryless, we demonstrate that the “time” component (how early or how late patients were in their disease progression) is actually encapsulated to some extent within out state definition. We conducted an analysis about the “time specificity” of each state (Figure 29). In the datasets, each set of data points is numbered for each patient, from the start until the end of the data collection. These represent time steps, and can take values from 1 up to 18 (corresponding to 72h of data collection). We computed the average value of the time steps in each state. While some states are only populated by patients at the early phase of their sepsis (left part of the plot), some exclusively contain patients at a late stage (right hand side of the plot). This demonstrates that the state definition captures to some extent the time elapsed since the sepsis onset or ICU admission.

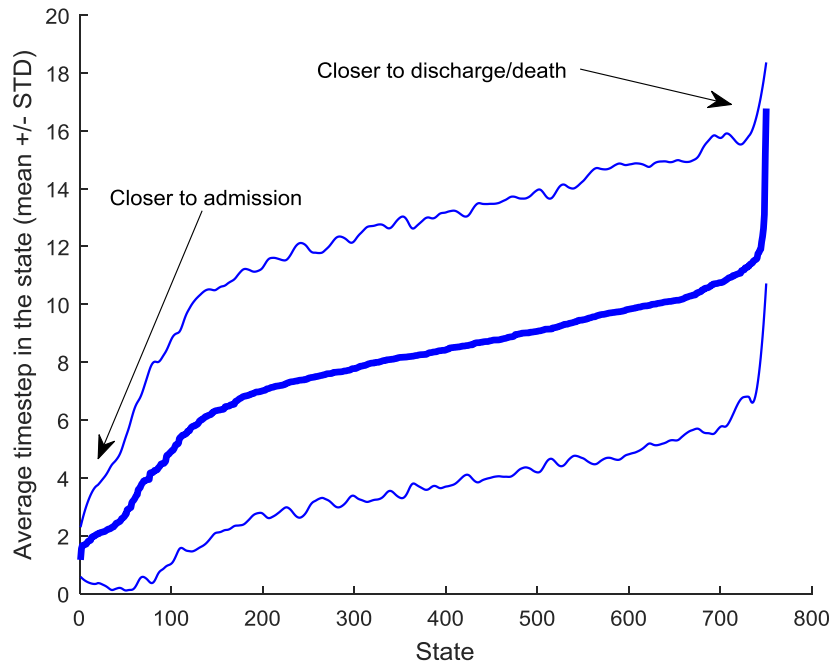


Figure 29: Average time step in each state, sorted by increasing values. Thin blue lines denote the standard deviation. While some states are only populated by patients at the early phase of their sepsis (closer to admission), some exclusively contain patients at a late stage (closer to discharge or death). This demonstrates that the state definition captures to some extent the time elapsed since the sepsis onset, despite the chain being mathematically memoryless.

Defining reward and penalty

The outcome that we intend to optimise is mortality (at 90 days or in hospital), so a reward was assigned to survival and a penalty to death. Each final transition of each trajectory leads to an absorbing state (death or discharge). Only these transitions are associated with a penalty of -100 points or a reward of +100 points, respectively. We did not assign any reward or penalty to the intermediate transitions to non-absorbing states. We could have assigned intermediate rewards to short-term resuscitation goals, similarly to the methodology used by Prasad et al. (Prasad et al., 2017). For example, we could have assigned a reward to a patient reaching a normal blood pressure, urine output, clearing his lactate, etc. The problem is that optimal targets for these parameters are not known at the patient level. For example, while it is important to avoid hypotension, there is still uncertainty about what blood pressure is optimal. Previous randomised controlled trials have shown no difference in mortality rates between different blood pressure targets in sepsis and it is likely that patients need individualised targets and that these might vary over time (Asfar et al., 2014; Lamontagne et al., 2016). There is also plenty of evidence that targeting or improving short-term physiological rewards (blood pressure (Asfar et al., 2014;

Lamontagne et al., 2016), urine output (Myburgh et al., 2012), oxygenation (ARDSNet, 2000)) can ultimately lead to worse long-term survival. We have therefore selected longer-term survival (90 days) as our reward signal as this is what matters to patients. The reinforcement learning approach allows us to assess and operate towards longer-term outcomes from a series of decisions, instead of just single action outcomes (as in causal modelling applications). Because 90-day mortality was not available in the eRI, hospital mortality was used as the outcome of interest. We verified first that the model performed well in the MIMIC-III database when training the model using hospital mortality and 90-day mortality (see Table 7, Table 8, Figure 34 and Figure 35). This sanity check supported the extension of the framework into the eRI data.

Patient trajectories

The trajectory of a patient can now be expressed as a n -by-3 array, with n rows of {state, action, reward} tuples. An example is shown in Table 6. It is also possible to plot the trajectories of patients onto a 2D or 3D projection of the state space, and to estimate their risk of mortality in real-time (Figure 30). On the figure, we display in the 2D state space the path of 2 separate patients as they evolve across distinct states, towards survival (green) or death (yellow). A more detailed analysis of what these patients may look like from a clinical standpoint is described in detail in chapter 5. We used a k-nearest neighbour method to estimate the risk of mortality of these patients at each time step (simply by computing the average mortality of his 30 nearest neighbours, which is a method described elsewhere (Sakr et al., 2017)). The cross-validated area under the receiver operator curve (AUROC) of this mortality prediction method is only fair, at around 0.75.

Step	State	Action	Reward
1	116	6	0
2	116	8	0
3	709	11	0
4	313	11	0
5	191	1	0
6	752	n/a	-100

Table 6: Example of a patient’s trajectory, defined as a {state, action, reward} array.

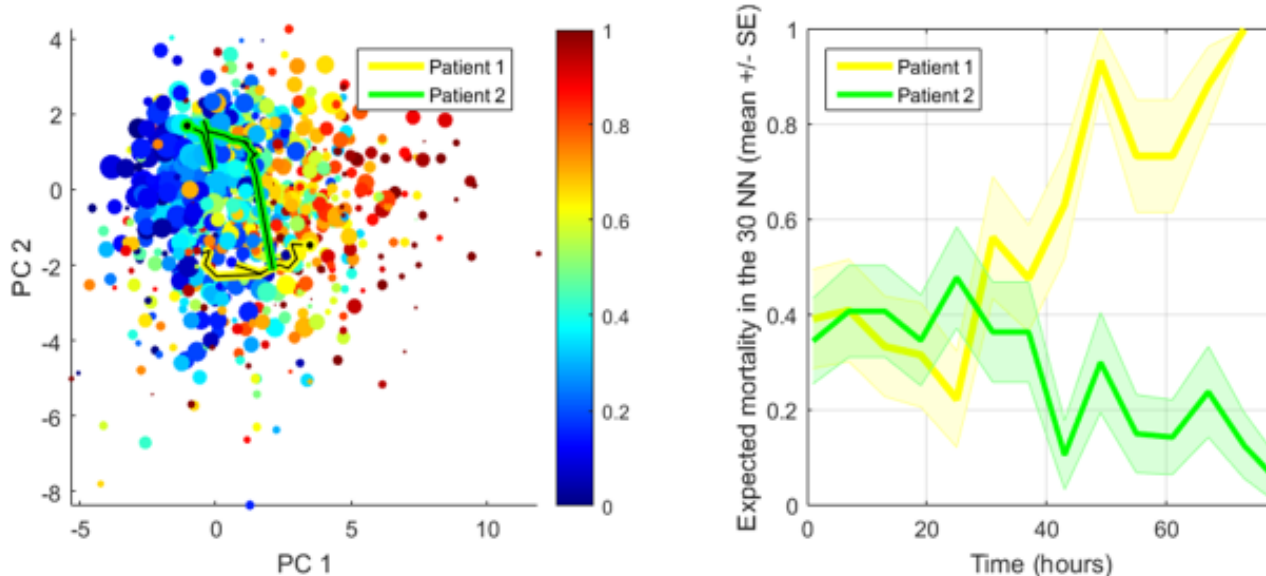


Figure 30: Displaying trajectories of patients in the state space. We show the path of 2 separate patients as they evolve across distinct states, towards survival (green) or death (yellow). On the right panel, the risk of mortality of these patients can be estimated continuously by a k-nearest neighbour method. The cross-validated AUROC of this mortality prediction method is around 0.75.

Discount factor gamma

The discount factor γ defines the horizon of the agent, which is how much importance is given to future rewards compared to the reward in the current state. γ can take values between 0 and 1, and needs to be set strictly more than 0 and less than 1 for the algorithm to converge (Sutton & Barto, 2018). We selected a value for gamma of 0.99, since we considered that immediate mortality was as bad as delayed mortality. Figure 31 shows the horizon of the RL agent for different values of gamma. We plot the Q values at all timesteps, expressed as a fraction of the initial value, following the formula γ^{timestep} . The figure shows that with values of gamma lower than 0.99, the discounting effect is important at 72 hours, leaving little value to delayed rewards. This is another argument for selecting a gamma of 0.99, since the median length of stay is around 3 days in our datasets.

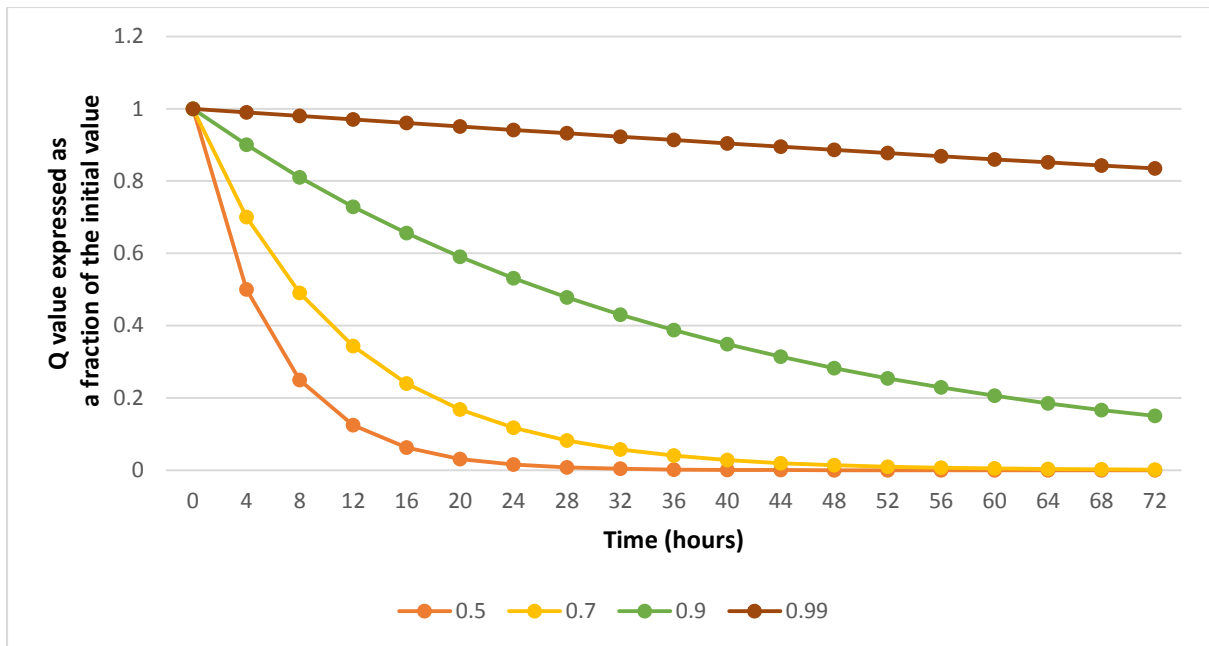


Figure 31: Horizon of the RL agent for different values of gamma. With values of gamma lower than 0.99, the discounting effect is important at 72 hours, leaving little importance to delayed rewards. The median length of stay in ICU being around 3 days in the datasets, we selected a gamma of 0.99.

RL algorithms

Once the MDP was built, we ran the two algorithms described in the background chapter:

- 1) TD-learning of the Q function was used to estimate the value of the clinicians' policy, by observing all the drug prescriptions in existing records and computing the average value of each treatment option, at the state level (Sutton & Barto, 2018). We stopped the evaluation after processing 500,000 patient trajectories with resampling, when the value of the estimated policy reached a plateau (Figure 32).
- 2) In-place policy iteration estimated the optimal policy. The goal of policy iteration was to identify the decisions that maximize the long-term sum of rewards, hence the expected survival of patients.

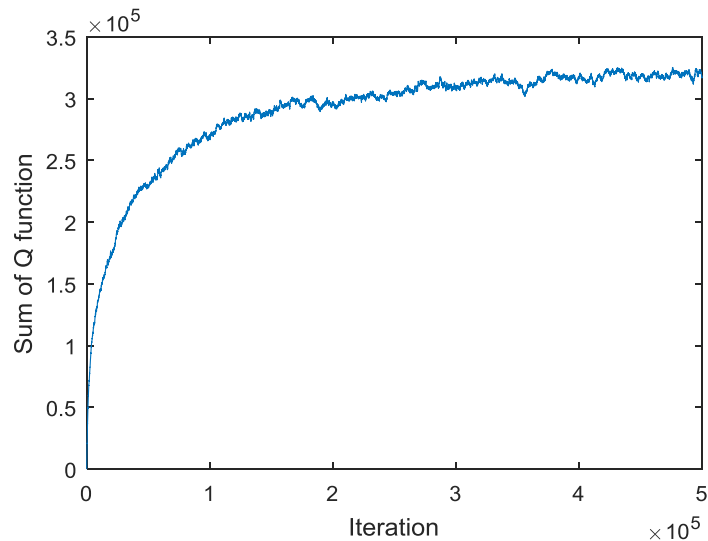


Figure 32: Sum of the Q function during 500,000 iterations of the TD-learning algorithm. Convergence of the algorithm is guaranteed when the Q function stabilises.

Discussion on model construction

The choice of which patient parameters were used to define patient health states can be discussed. We elected to have very little assumptions about what data would be relevant to the question at stake (the prescription of an optimised dose of intravenous fluids and vasopressors), and fed into the training dataset much of the (mostly numerical) data that was available from patients in the MIMIC-III set. The upside of this strategy is that we limit our assumptions of what features are relevant or not. The disadvantage is that signal may be diluted in noise, in particular at the step of the state definition by clustering. Indeed, if we use the Euclidean distance between vectors of normalised patient features, then all the features have as much importance in defining cluster membership. While this is expected from a mathematical standpoint, a clinician would argue that mean arterial pressure or dose of noradrenaline have in general more importance in determining “how sick” a patient is than -say- their magnesium level or their prothrombin time. We demonstrate when evaluating individual patient trajectories that state definition could sometimes be improved, for example in large states that contain an array of patients of various severity.

Future directions for model development include feature selection and feature engineering, for example by rescaling the mean arterial pressure so more classes can be defined in the clinically relevant range (e.g. between 60 and 80 mmHg). Along the same lines, some features that are currently present in the model are highly collinear (e.g. systolic, diastolic and mean blood pressure), or even redundant (PaO_2 , FiO_2 and $\text{PaO}_2/\text{FiO}_2$). We discuss in the final chapter in this thesis how the clustering could be improved

(Chapter 6, “Future directions for model development”). Further experiments to determine whether they are important to the model performance are warranted.

Chapter summary

In this chapter, we discussed how a simplified MDP model was built from MIMIC-III data to model the trajectories of patients with sepsis. Next, the algorithms that we described in the background chapter were used to evaluate the value of clinicians’ decisions (π_b) and to estimate an optimal policy (π_e). In the next chapter, we will discuss how the model built was evaluated and how we selected the most optimal policy among a set of candidate policies.

Chapter 4: Model evaluation and policy selection

In this chapter, we discuss how the model built (using the MIMIC-III training set) was evaluated and how we selected the most optimal policy among a set of candidate policies. All these analyses were carried out on the MIMIC-III data, so all the results and figures shown in this chapter apply to MIMIC-III only.

We have discussed in the introduction the challenges associated with the retrospective validation of a learning algorithm. We defined off-policy policy evaluation (OPE) as the task of predicting the performance of a newly generated RL policy (π_e , the evaluation policy, the AI policy) given historical data that has been generated by a different policy (π_b , the behaviour policy, the clinicians' policy).

Determining optimal actions for test records

The validation of the algorithms uses test records (from MIMIC-III test set, eRI and CCHIC) that were not exploited during the building of the model. The MIMIC-III test records have the same features than the MIMIC-III training samples, in the same units of measurement. The eRI and CCHIC records match many, but not all, of the MIMIC-III features (refer to Table 2), in the same units of measurement.

The optimal dose range of the two drugs according to the AI policy only depends on the current state the patient is in, since the AI policy π^* corresponds to a $\{750 \times 1\}$ vector controlling which action should be taken in every state of the MDP. We determined the cluster membership (or state membership) of these test samples according to whichever cluster centroid they fell closest to (using the Euclidean distance), after transformation or normalisation of the features, as appropriate (see illustration Figure 33).

Instead of using the cluster centroid method, we also tested matching a test sample t to 5 or 10 samples from the training set, and using their recommended dose of drugs (the mean or median dose among these 5 or 10 samples) as the optimal dose for the test sample t . In the end, for simplicity and computational efficiency, we only used the cluster centroid method.

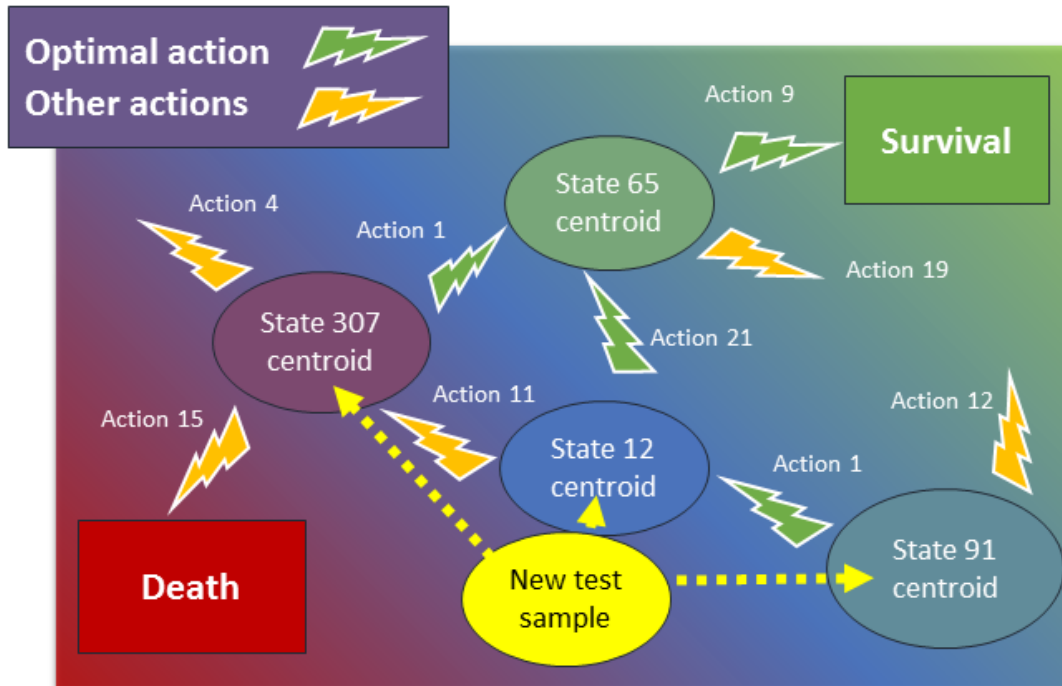


Figure 33: Illustration of the method used to determine the optimal action for a new test sample. The state membership of a new test sample (in yellow in this example) is not readily known, since this sample was not used during clustering. The state membership is determined according to whichever cluster centroid they fall closest to (state 12 here). Action 21 being the optimal action for state 12, it becomes the recommended action for the new test sample.

Model-based off-policy evaluation

Here, we fitted an MDP model from data via regression, and evaluate the policy against the model (Jiang & Li, 2015). We have discussed the limitations of such an approach above: *“Such a regression based approach has a relatively low variance and works well when the model can be learned to satisfactory accuracy. However, for complex real-world problems, it is often hard to specify a function class in regression that is efficiently learnable with limited data while at the same time has a small approximation error. Furthermore, it is in general impossible to estimate the approximation error of a function class, resulting in a bias that cannot be easily quantified”* (Jiang & Li, 2015).

We present here the results of this analysis for the MIMIC-III test set. Figure 34 represents the distribution of the value of the clinicians’ (evaluated by TD learning) and the AI policies (evaluated by policy iteration) in MIMIC-III test set (models optimizing hospital and 90-day mortality), using bootstrapping with 2,000 resamplings. Table 7 shows the estimated values of the policies, in both

models. According to these estimates, the value of the AI policy in the model-based evaluation was always higher than the clinicians'. The results of this analysis carried out in the eRI cohort are shown in the next chapter, where we also compare the value of the clinicians' and the AI policies.

	Clinicians' policy value	Model-based AI policy value
Hospital mortality	73 (72.4-73.8)	88.3 (88-88.6)
90-day mortality	53.7 (53-55.2)	82.3 (81.8-82.7)

Table 7: Estimated clinicians' and model-based AI policy value in the MIMIC-III test set. The numbers represent the median and interquartile range of the policy values.

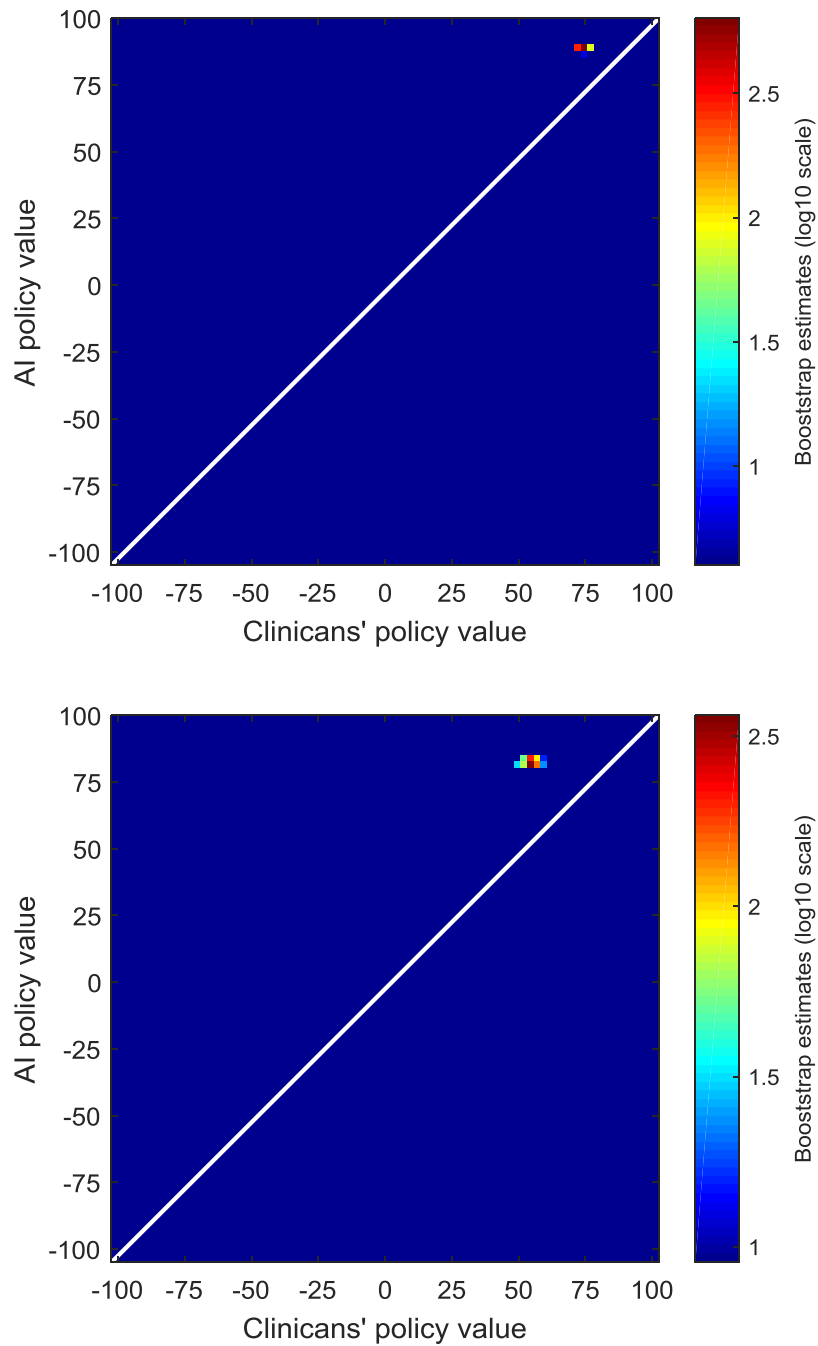


Figure 34: Model-based OPE in MIMIC-III test set, for models optimising hospital (top) and 90-day (bottom) mortality. Built by bootstrapping with 2,000 resamplings. Any point sitting above the diagonal line represents an instance where the AI agent identifies a decision that has a higher value than the clinicians'. According to these estimates, the value of the AI policy in the model-based evaluation was always higher than the clinicians'.

Importance sampling-based off-policy evaluation

Policy value estimators based on importance sampling (IS) avoid many of the limitations of model-based estimators. The key principle of IS is to correct for the discrepancy between the behaviour (π_b) and the evaluation (π_e) policies when learning from off-policy returns (Jiang & Li, 2015; Munos, Stepleton, Harutyunyan, & Bellemare, 2016; Precup et al., 2000). The correction uses the product of the likelihood ratios between π_b and π_e to produce an unbiased estimator of $V(\pi_e)$ (P. Thomas, Theodorou, & Ghavamzadeh, 2015). Hanna commented that weighted IS estimators are likely to underestimate the true value of π_e (since most importance weights are close to zero, the IS estimate will be pulled towards zero), which is preferable to overestimation in applications where safety is critical (Hanna et al., 2016).

Two problems arise when applying IS to off-policy evaluation (Hanna et al., 2016; Munos et al., 2016):

- 1) IS estimates can suffer from very large variance (mainly due to the variance of the product $\pi_e(a_1|x_1)/\pi_b(a_1|x_1) \cdots \pi_e(a_t|x_t)/\pi_b(a_t|x_t)$), which has motivated different variance reduction methods, in particular the use of weighted estimators. Therefore, we implemented a weighted version of IS. WIS is statistically consistent (i.e., $\text{WIS}(D) \rightarrow V(\pi_e)$ as $n \rightarrow \infty$). Clipping the IS estimates so they do not exceed pre-defined values has been proposed in the literature (Gottesman et al., 2019).
- 2) IS requires π_e to be soft. If π_e is deterministic, we can only estimate traces where π_e and π_b match. As soon as they diverge, the importance weight becomes zero, so does the IS estimator, causing many traces to be cut prematurely and blocking learning from full returns. To counter this limitation, we softened π_e , which now recommends taking the suggested action 99% of the time, and any of the other actions a total of 1% of the time. This allows assessing the entirety of the patient trajectories. Thomas proposed using mixed policies to address this problem (P. Thomas et al., 2015). In his approach, he defines a mixed policy $\mu_{\alpha, \pi_1, \pi_2}$ to be a mixture of π_1 and π_2 with mixing parameter $\alpha \in [0, 1]$. As α increases, the mixed policy becomes more like π_2 , and as α decreases it becomes more like π_1 . Formally, $\mu_{\alpha, \pi_1, \pi_2}(a|s) = \alpha\pi_1(a|s) + (1 - \alpha)\pi_2(a|s)$.

Our goal is to estimate the value of π_e from data trajectories. We define $\rho_t := \pi_e(a_t|s_t)/\pi_b(a_t|s_t)$ as the per-step importance ratio, $\rho_{1:t} := \prod_{t'=1}^t \rho_{t'}$ as the cumulative importance ratio up to step t , $w_t = \sum_{i=1}^{|D|} \rho_{1:t}^{(i)}/|D|$ as the average cumulative importance ratio at horizon t in dataset D , $|D|$ the number of trajectories in D (Jiang & Li, 2015; P. S. Thomas et al., 2015).

The trajectory-wise WIS estimator is given by:

$$V_{WIS} = \frac{\rho^{1:H}}{w_H} (\sum_{t=1}^H \gamma^{t-1} r_t)$$

Then, the WIS estimator is the average estimate over all trajectories, namely:

$$WIS = \frac{1}{|D|} \sum_{i=1}^{|D|} V_{WIS}^{(i)}$$

Where $V_{WIS}^{(i)}$ is WIS applied to the i -th trajectory.

The next step is to obtain confidence intervals on the IS estimates. Formally, given a set of n trajectories, $D = \{H_1, \dots, H_n\}$, where $H_i \sim \pi_b$ for some π_b , a π_e , and a confidence level, $\delta \in [0, 1]$, we want to approximate a confidence lower bound, $V\delta(\pi_e)$, on $V(\pi_e)$ such that $V\delta(\pi_e) \leq V(\pi_e)$ with probability of at least $1 - \delta$.

Two main approaches have been proposed to generate confidence bounds on the IS estimates: concentration inequality and bootstrapping (Hanna et al., 2016; Precup et al., 2000; P. S. Thomas et al., 2015; P. Thomas et al., 2015). Concentration inequality is a statistical method that provides exact probability bounds on how a random variable deviates from its expectation, but often require an impractical amount of data before the bounds are tight enough to be useful (P. S. Thomas et al., 2015). Bootstrapping, on the other hand, provides approximate but accurate confidence bounds, with less data than exact HCOPE methods, and is safe enough in high-risk applications such as healthcare (Algorithm 3) (Hanna et al., 2016; P. Thomas et al., 2015). Hanna also confirmed that IS- or PDIS-based bootstrapping provides the safest approach for HCOPE when the model estimation error is high, which we may take as an assumption to maximise the safety of our algorithm (Hanna et al., 2016). Finally, note that, since we only assign rewards to the final transition in each trajectory, WIS and PDWIS are equivalent in our application (Jagannatha, Thomas, & Yu, 2018).

Algorithm 1 Bootstrap Confidence Interval

Input is an evaluation policy π_e , a data set of trajectories, \mathcal{D} , a confidence level, $\delta \in [0, 1]$, and the required number of bootstrap estimates, B .

input $\pi_e, \mathcal{D}, \pi_b, \delta, B$

output $1 - \delta$ confidence lower bound on $V(\pi_e)$.

1: **for all** $i \in [1, B]$ **do**

2: $\tilde{\mathcal{D}}_i \leftarrow \{H_1^i, \dots, H_n^i\}$ where $H_j^i \sim \mathcal{U}(\mathcal{D})$ // where \mathcal{U} is the uniform distribution

3: $\hat{V}_i \leftarrow \text{Off-PolicyEstimate}(\pi_e, \tilde{\mathcal{D}}_i, \pi_b)$

4: **end for**

5: $\text{sort}(\{\hat{V}_i | i \in [1, B]\})$ // Sort ascending

6: $l \leftarrow \lfloor \delta B \rfloor$

7: **Return** \hat{V}_l

Algorithm 3: Pseudocode for bootstrap confidence interval, from Hanna (Hanna et al., 2016).

Figure 35 and Table 8 show the result of this analysis for the MIMIC-III test set, for models optimising hospital and 90-day mortality. The median clinicians' policy value was estimated at 73.8 (IQR 73-74.1) and 51.9 (IQR 50.7-53.1) for models optimising hospital and 90-day mortality, respectively. In comparison, the WIS-based estimator of the AI policy gave values of 87.6 (IQR 87.1-88.4) and 87.7 (IQR 85.2-88.9) for models optimising hospital and 90-day mortality, respectively. It is interesting to note that the actions of clinicians had a lower value when trying to optimise 90-day mortality than hospital mortality, as if humans were better at optimising short-term than long-term outcomes. The estimates of the AI policy did not elicit this trend. The results for the eRI cohort are shown in the next chapter.

Model optimising:	Clinicians' policy value	WIS-based AI policy value
Hospital mortality	73.8 (73-74.1)	87.6 (87.1-88.4)
90-day mortality	51.9 (50.7-53.1)	87.7 (85.2-88.9)

Table 8: Estimated clinicians' and WIS-based AI policy value in the MIMIC-III test set. The numbers represent the median and interquartile range of the policy values. Built by bootstrapping with 2,000 resamplings.

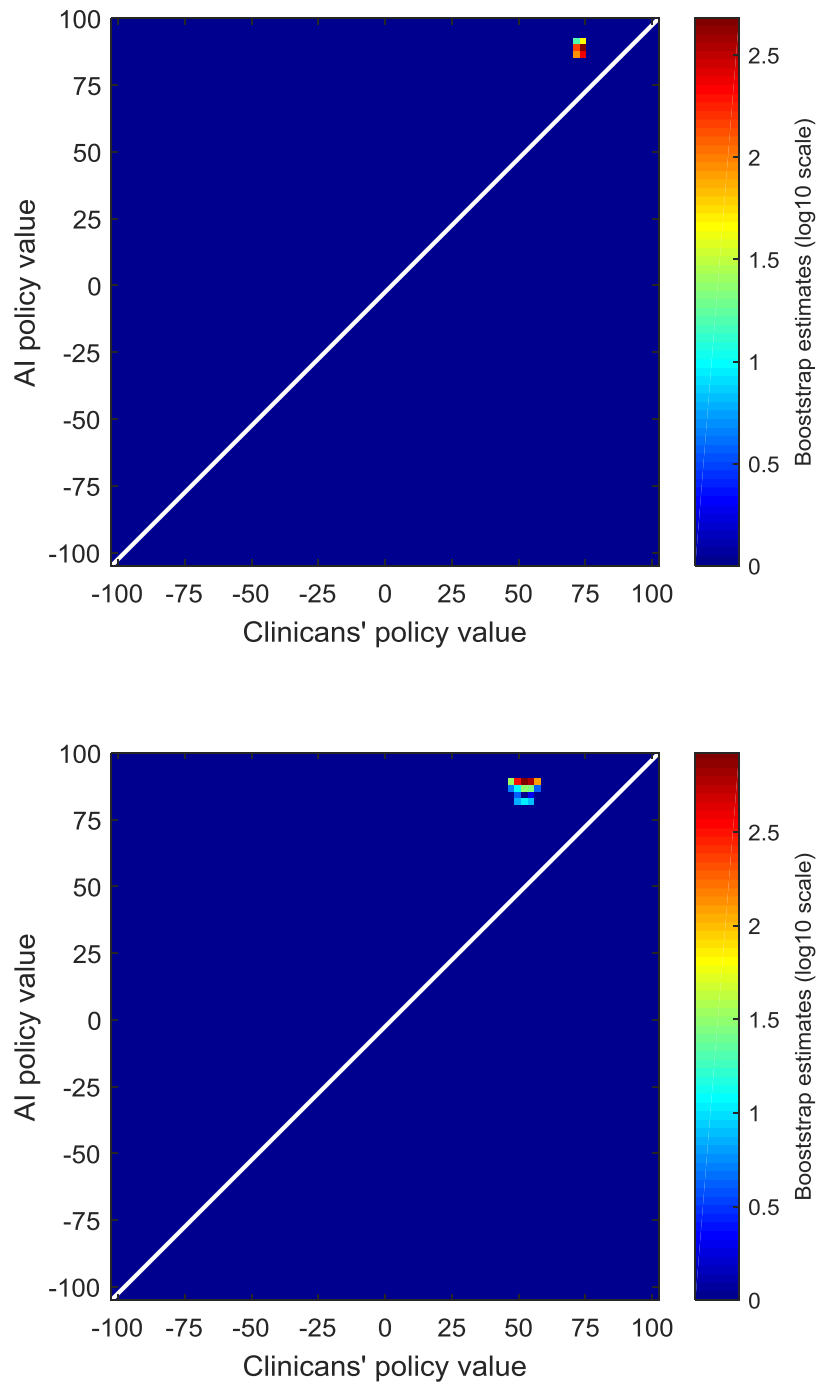


Figure 35: IS-based OPE in MIMIC-III test set, for models optimising hospital (top) and 90-day (bottom) mortality. Built by bootstrapping with 2,000 resamplings. According to these estimates, the value of the AI policy in the IS-based evaluation was always higher than the clinicians', since all data points are above the white diagonal line.

Selecting the best model using high confidence off-policy evaluation (HCOPE)

In this research, it was crucial to obtain reliable estimates of the performance of this new policy without deploying it, since executing a bad policy would be dangerous for patients (Hanna et al., 2016; P. S. Thomas et al., 2015). We used bootstrapping to estimate the true distribution of the WIS policy value estimates in the MIMIC-III 20% validation set. We built 500 different models using 500 different clustering solutions of the training data, and the selected final model maximised the 95% confidence lower bound of the AI policy (Hanna et al., 2016). This relates to the concept of safe RL, where the policy that has the highest lower confidence bound is selected, otherwise we hold on to the current behaviour policy if none of the bounds is better than the behaviour policy's value (Jiang & Li, 2015; P. S. Thomas et al., 2015; P. Thomas et al., 2015). Figure 36 (left) shows that this bound consistently exceeded the 95% confidence upper bound of the clinicians' policy, provided that enough models were built. The figure shows the highest 95% confidence upper bound of the clinicians' policy, along with the 95% lower bound of the best AI policy, during the process of building the 500 models. The selected final model is the one that achieves the highest 95% confidence lower bound value of the WIS estimate among the 500 candidates. The right plot shows the distribution of the estimated value of the clinicians' policy, the AI policy (with the chosen final model), a random policy and a zero-drug policy across the 500 models, in the MIMIC-III test set. The AI policy was then tested on the independent eRI and CCHIC datasets (these results are presented in the next chapter).

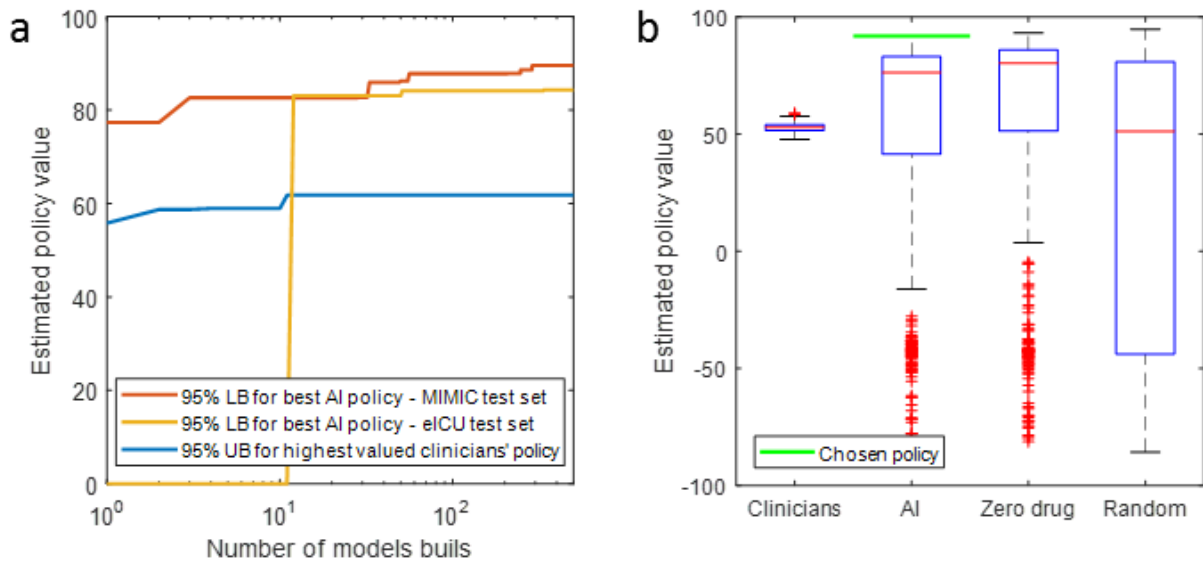


Figure 36: Selection of the best AI policy. a, evolution of the 95% lower bound (LB) of the best AI policy and 95% upper bound (UB) of highest valued clinicians' policy during the building of 500 models. After only a few models, a higher value for the AI policy than the clinicians' practice, within the accepted risk, is guaranteed. b, distribution of the estimated value of the clinicians' practice, the AI policy, a random policy and a zero-drug policy across the 500 models, in the MIMIC-III test set. The chosen AI policy maximises the 95% confidence lower bound.

Other policy evaluation methods

Visual methods

These methods provide a simple description of the model's behaviour and features but they provide no statistical guarantee of the validity of the model (P. S. Thomas & Brunskill, 2016). Caution must be exerted with these methods, since even if a model estimates perfectly the performance of the policy used to generate the data from which the model was trained, the model may still be a very poor representation of the reality. For example, Thomas and Brunskill have demonstrated that some models can appear to be perfect (in terms of predicting the performance of the policy that was used to generate the data from which the model was trained) when in reality they are very weak models (P. S. Thomas & Brunskill, 2016).

Nevertheless, apparent good model calibration was shown by plotting the relationship between the state-action value of clinicians' actual policy and patients' 90-day mortality, in the development cohort

(Figure 37, left). In Figure 37 (right), we show the average return measured in survivors and non-survivors.

Among visual methods, we can also observe the distribution of doses of both drugs according to the clinicians and the AI policy (results for the test sets are shown in the next chapter). This will provide high-level information about treatment patterns, such as which proportion of patients received vasopressors versus what this number would be if the AI policy was followed, or the proportion of patients on high doses of fluids according to both policies, for example.

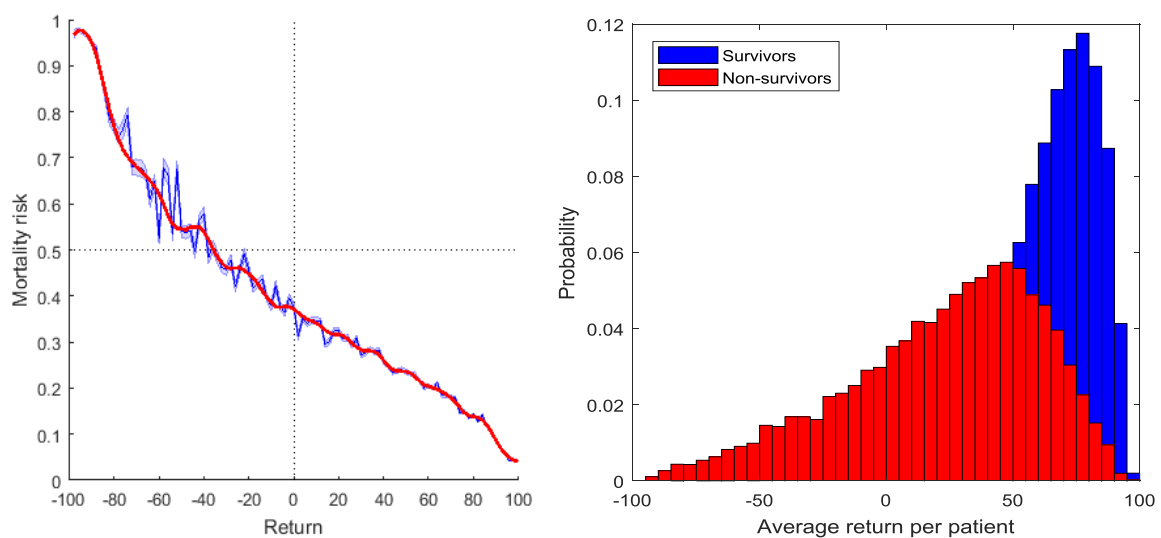


Figure 37: Model calibration. Left: relationship between the return of clinicians' decisions and patient 90-day mortality, in the MIMIC-III training set (n=13,666 patients). Return of actions were sorted into 100 bins and the mean observed mortality (blue line for raw, red line for smoothed) was computed in each of these bins. Decisions with a low return were associated with a high risk of mortality, while choices with a high return led to better survival rates. Right: average return in survivors and non-survivors in the MIMIC-III training set. Both figures were generated by bootstrapping in the training data with 2,000 resamplings.

Direct policy assessment methods

Relationship between doses received and mortality

Validating a suggested dose of drugs in a retrospective analysis is difficult. Obviously, it is impossible to know the outcome of an action that was never taken. Therefore, we analysed patient mortality when the dose actually administered corresponded to or differed from the dose suggested by the AI agent. We hypothesised that it is possible to get retrospective evidence for the model validity if we show that the patients who followed the AI policy had the best outcome, and that giving a different dose than the one suggested by the AI policy was associated with increasing mortality rates. As described in the introduction chapter, this is somewhat similar to what Prasad et al. did in their research on applying RL to weaning of mechanical ventilation (Figure 11) (Prasad et al., 2017).

We computed the difference between doses given to the patient and doses recommended by the AI agent, which we refer to as the “dose excess”. Figure 38 shows the dose excess in the MIMIC-III test set, for a model optimising hospital mortality. With this approach, we confirmed that the patients who received doses similar to the doses recommended by the AI agent had the lowest mortality, and that giving more or less than the AI policy of either treatment was associated with increasing mortality rates, in a dose-dependent fashion.

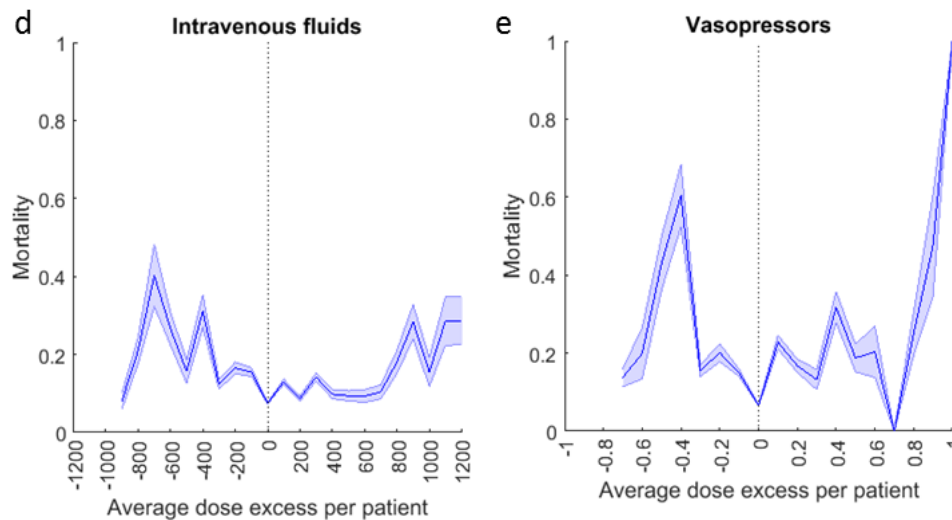


Figure 38: Internal validation in the MIMIC-III test set: model optimizing hospital mortality. Average dose excess received per patient of intravenous fluids (left) and vasopressors (right), and corresponding mortality. The figure is built by bootstrapping with 2,000 resamplings. The lowest expected mortality was found when the dose actually administered to the patients matched the dose suggested by the AI policy. The shaded area represents the standard error of the mean.

Importance of choosing the most optimal action

In each state, we can rank all 25 actions according to their state-action $Q(s, a)$ value, and define the most optimal action (in terms of impact on the patient’s risk of mortality), the second most optimal action, and so on until the least optimal (the worst) action. In parallel, we know the outcome of the patients for any given $\{s, a\}$ pair. Combining these two data, we can compute patient mortality for the all 25 actions averaged over all 750 states, when the actions are ranked in terms of their “optimality”. Figure 39 shows this analysis and confirms that choosing even the second best option is associated with worse outcomes.

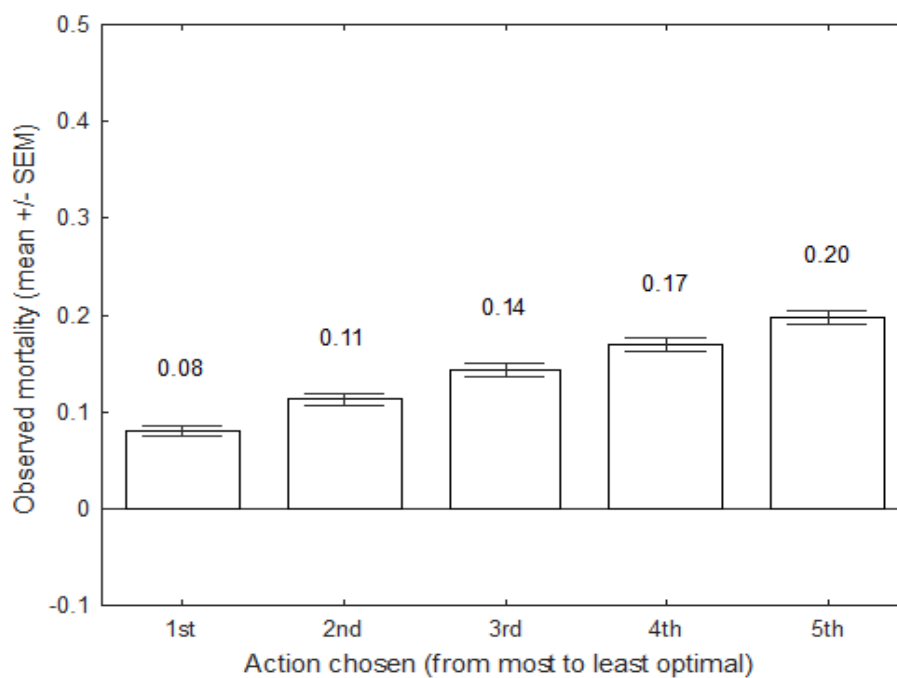


Figure 39: Importance of choosing the most optimal action. The figure shows the observed patient mortality in MIMIC-III averaged across the 750 states, depending on whether they took the 1st, 2nd, etc. most optimal action. Choosing even the 2nd best option is associated with worse outcomes.

Mortality at the state level for actual and suggested actions

To estimate the outcomes associated with following various policies, one possibility is use simulation and use our model to generate a large number of artificial patient trajectories under various policies. This method broadly falls under the spectrum of Monte Carlo simulations (Bonate, 2001). Monte Carlo simulations in general produce distributions of possible outcome values from repeating over and over again calculations of a desired result, each time using different set of random values for input factors,

sampled from a given probability distribution. We applied this method by generating artificial patient trajectories from any given state until death or discharge, when following various policies within the constraints of our model, defined by its transition probabilities.

In this Monte Carlo simulation analysis performed on the MIMIC-III training data, we estimated the average mortality across states by following 3 different strategies: 1) all the actions are sampled from the actual physicians' policy ; 2) the initial action is the recommended action by the AI agent, then all subsequent actions are sampled from the physicians' policy; and 3) the initial action is the most frequent action taken by physicians, then all subsequent actions are sampled from the physicians' policy. 2,000 Monte Carlo draws were used in each scenario. The mortality averaged over all 750 states is shown in Figure 40. The estimated mortality using this approach was 19.37% (SEM 0.86%), 17.87% (SEM 0.84%) and 19.27% (SEM 0.86%) in groups 1, 2 and 3, respectively. Of note, the 19.37% mortality in scenario 1 differs from the 22.5% actual mortality seen in the cohort because we did not average the mortality across states based on the initial state distribution (no weighted average) but simply averaged across the 750 states. The demonstration of accuracy of this method is shown in Figure 50. A one-way ANOVA analysis confirmed that the mortality in the second group (where the only difference is that the initial action taken is the action suggested by the AI agent) was significantly lower than the two others ($p < 1e-5$). There was no difference in estimated mortality between patients in groups 1 and 3 ($p > 0.05$): the estimated mortality was similar whether using the full scope of clinicians' actions or simply the mode of their treatment strategy.

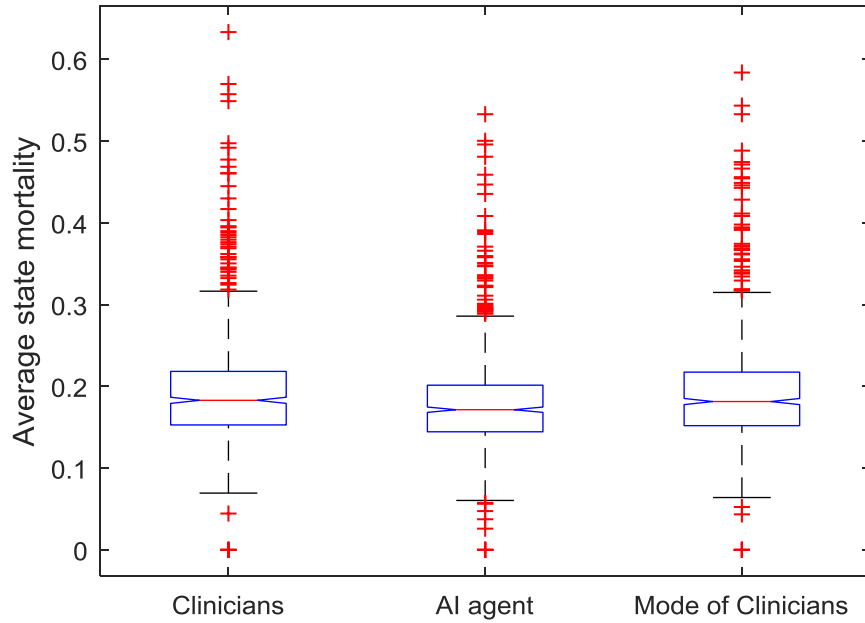


Figure 40: Monte Carlo simulations of the mortality at the state level when following actual clinicians’ policy (group “Clinicians”), the suggested optimal action at the first step, then the clinicians’ policy (group “AI agent”) or the most frequent action chosen by clinicians at the first step, then the clinicians’ policy (group “Mode of Clinicians”). The estimated mortality in the second group was significantly lower than the two others (1-way ANOVA $F = 11.41$, $p < 1e-5$) and there was no difference between groups 1 and 3 (1-way ANOVA $p > 0.05$).

Next, we used a similar method to build Kaplan-Meier survival curves from any given state of the system (see examples Figure 41). We blended policies so actions would be sampled either from the clinicians’ or the AI policy, with a 50% probability. We compare this to estimating patient survival when following exclusively the clinicians’ policy. Sampling actions from the AI policy results in a significant improvement in estimated survival.

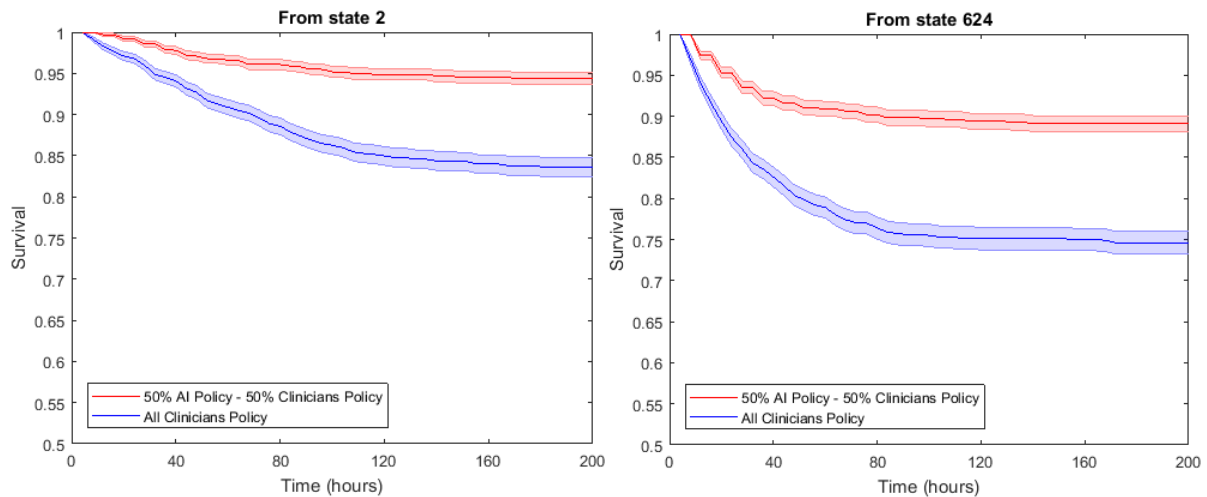


Figure 41: Examples of Kaplan-Meier survival curves from state 2 and 624, when following either the clinicians policy (in blue) or a balanced blend of clinicians and the AI policy (in red).

Model interpretability: the “right to explanation”

Interpretability – or explainability – is a key aspect of machine learning developments, in particular in high-risk environments (Doshi-Velez & Kim, 2017; Goodman & Flaxman, 2016; Miller, 2017; Narayanan et al., 2018). In healthcare applications, all stakeholders (patients, clinicians, policy makers) are entitled to ask for explanations as to why an AI algorithm is reaching a certain decision.

The difficulty arises from the fact that interpretability, unlike performance criteria such as prediction accuracy, cannot be easily quantified (Doshi-Velez & Kim, 2017). The challenge of model interpretability has motivated updates in international regulations such as the General Data Protection Regulation (GDPR) which was adopted by European Parliament in April 2016 and became enforceable throughout the European Union in May 2018 (<https://www.eugdpr.org>). This law intends to unify regulation on data protection and privacy within the EU. It restricts automated individual decision-making which significantly affects users and introduces a “right to explanation” whereby a user can ask for a justification of an algorithmic decision that was made about them (Goodman & Flaxman, 2016). The ‘right to explanation’ is a key part of achieving accountability, however, ensuring algorithmic fairness and explainable AI will clearly represent a challenge for both academia and the industry.

Doshi-Velez and Kim have classified methods to improve model interpretability into 2 categories: usefulness and via a quantifiable proxy (Doshi-Velez & Kim, 2017). The first argument is that if a system is useful for an application, then it must be somehow interpretable. This can be related to concepts introduced by Miller and inspired from the social sciences and psychology literature. Miller

(and others) highlighted how humans, unlike machines, also care about simplicity, generality, coherence of the explanations and usefulness of tools (Miller, 2017; Narayanan et al., 2018). The second type of interpretability methods relies on the use of a quantifiable proxy. The intuition here is that researchers may first demonstrate the general interpretability of some core method (e.g. a particular classification or regression method), then apply it for a specific application (e.g. predict the onset of acute kidney injury) (Doshi-Velez & Kim, 2017).

Gaining an insight into the model representation underlying the model decisions can be made, for example, by looking at which variables motivate the strategy of the learnt policy. We can attempt to do this in the two test datasets by estimating the relative feature importance for predicting the administration of the medications (as a binary feature: drug on or off) using a random forest classification model, and compare results for clinicians to the AI policy (Figure 61 for eRI and Figure 64 for CCHIC). This confirmed that the decisions suggested by the RL algorithm were clinically interpretable and relied primarily on sensible clinical and biological parameters, such as serum lactate levels, mean blood pressure, or urine output.

Another method to try to improve explainability of the AI policy decisions relies on the use of a patient similarity metric. We can look at the treatment given in – for example – 500 patients similar to a test record and compare it to the treatment suggested by the AI policy. This relies on 2 assumptions: 1) similar patients need similar doses of drugs and 2) there exists at any time an optimal dose range associated with the best survival. Model validation implies that survival is the highest in patients who received a dose similar to the recommended dose. Figure 42 illustrates this concept with a cherry-picked example of what this result may look like in a given state, for both intravenous fluids and vasopressors. The dose of drugs received by the 500 nearest neighbours are divided into 10 deciles, and the average mortality for each decile was computed. In some states, there seems to be an optimal dose of drug associated with the highest survival, and a dose-effect relationship between dose and mortality is observed. Showing that the dose suggested by the AI agent is associated with good survival in patients similar to a new “target” patient could increase the confidence of clinicians in the accuracy and safety of the AI agent. A key limitation of this approach is that we only assess the value of the one-step current decision, instead of the value of the full (discounted) sequence of decisions (which is assessed with the policy value evaluation methods).

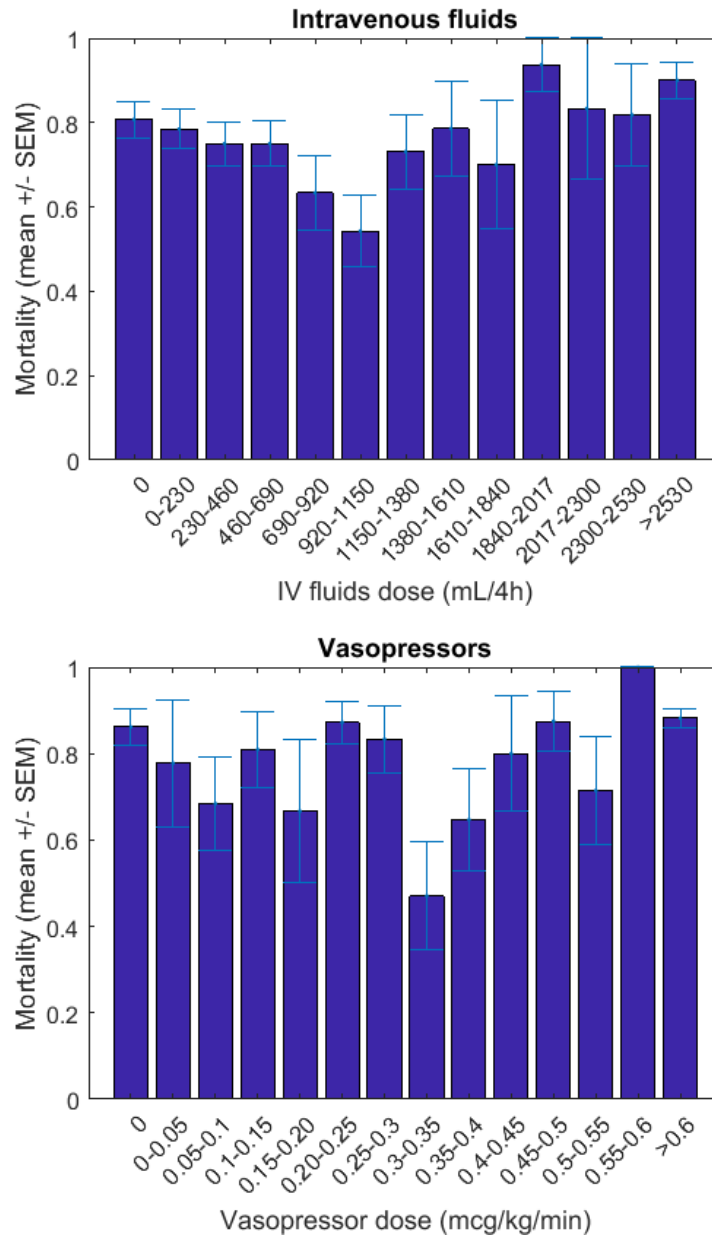


Figure 42: Concept of improving model explainability using a patient similarity metric. Here, we identified 500 patients who are similar to a target patient, analysed what doses of drugs were given to these patients and what was their mortality rate. In some states, there seems to be an optimal dose of drug associated with the highest survival, and a dose-effect relationship between dose and mortality for both intravenous fluids and vasopressors.

Further results

Variability of AI policy in different models

Different clustering solutions produce different optimal policies. To demonstrate this, we plot in Figure 43 the distribution of optimal suggested decisions (over the 750 states) in 5 randomly selected models, coming from the 500 candidate models. The optimal treatment strategy depends on how the clinical states of the MDP are defined.

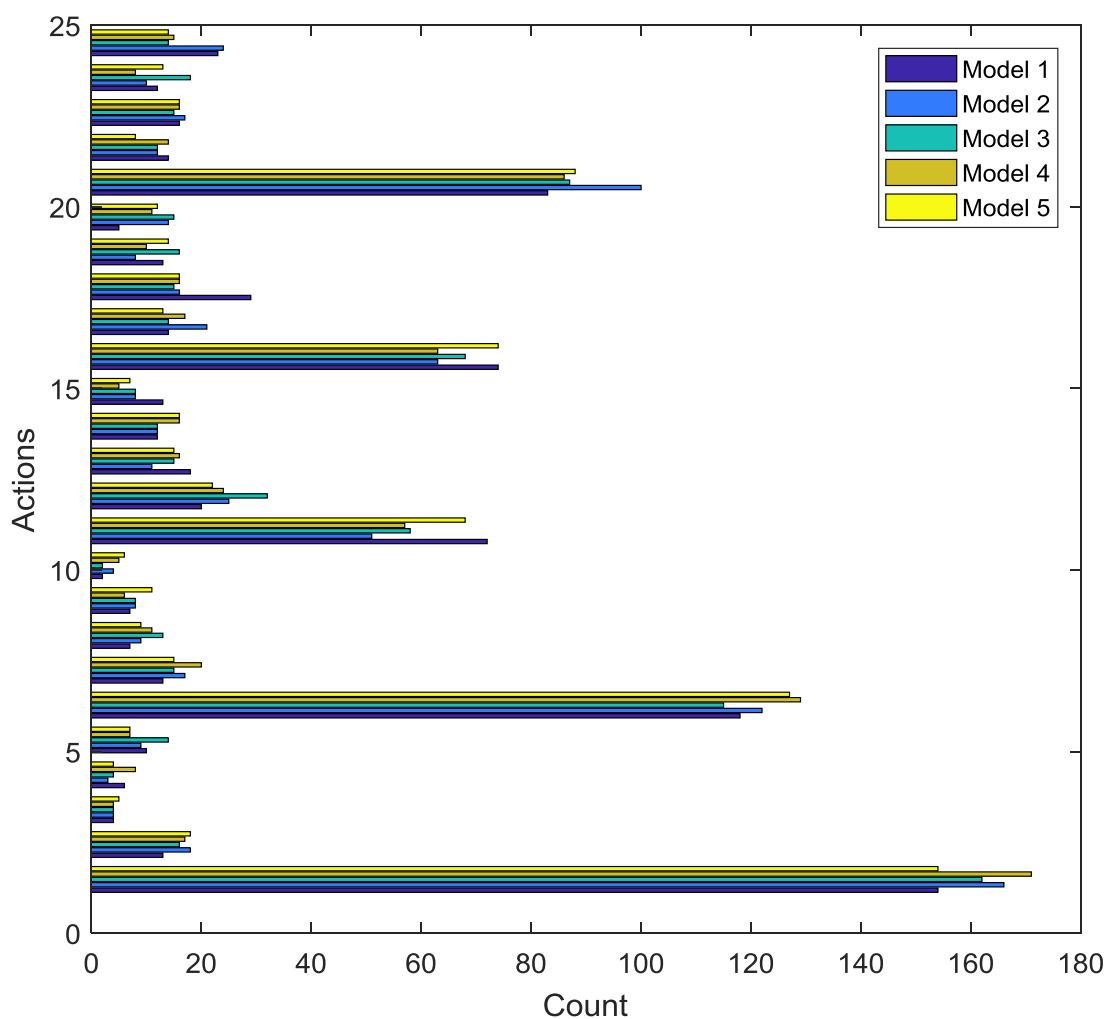


Figure 43: Distribution of optimal suggested decisions in 5 randomly selected models, coming from the 500 candidate models. Each clustering solution leads to a different distribution of optimal actions. The optimal treatment strategy depends on how the clinical states are defined.

Qualitative comparison of the clinicians' and AI policies: "internal validation"

Figure 45 shows the distribution of treatment doses (the distribution of actions) according to clinicians' and AI policies, in the MIMIC-III test set, in a model optimising hospital mortality. This can be compared to Figure 44 developed on the eRI data ("external validation"). On average, the AI agent recommended lower doses of intravenous fluids and higher doses of vasopressors, compared to clinicians' actual treatments.

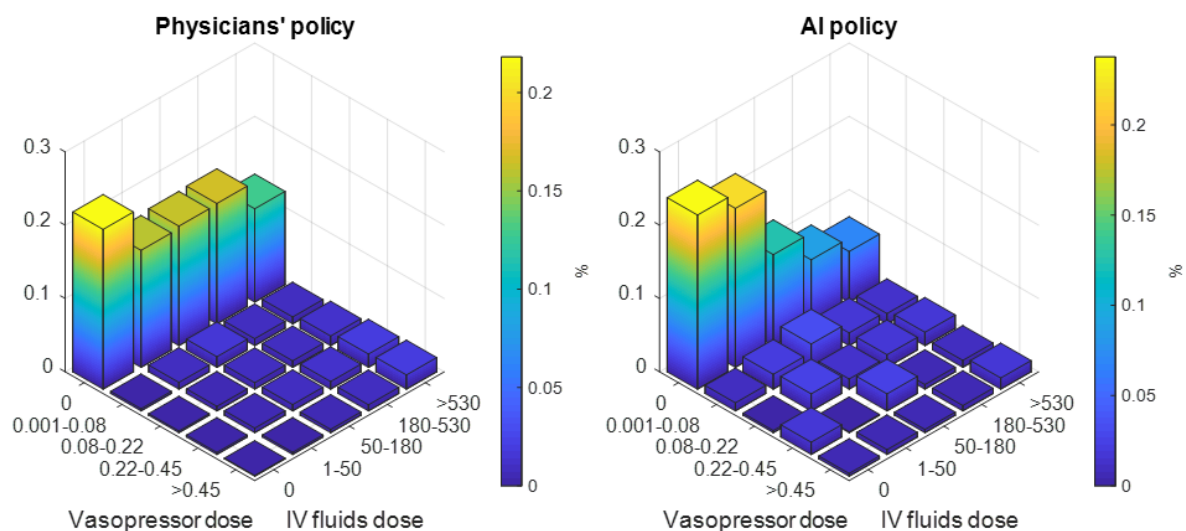


Figure 45: Internal validation in the MIMIC-III test set: model optimizing hospital mortality. Visualization of the clinicians' and AI policies. All actions were aggregated over all time steps for the 5 dose bins of both medications. On average, patients were administered more intravenous fluid and less vasopressor medications than recommended by the AI policy. Vasopressor dose is in mcg/kg/min of norepinephrine equivalent and intravenous fluids dose is in mL/4 hours.

Next, we compared the doses given to patients to doses recommended by the AI agent, depending on the current SOFA score of each record, in the MIMIC-III training set (in order to have a large record sample) (Figure 46). Similarly to physicians, the AI agent recommends increasing doses with increasing severity. While patients with a low severity did not receive vasopressors, the AI agent, on average, recommends administering them at a low dose. The AI agent recommends lower doses of IV fluids than clinicians for patients with an average-to-high severity (SOFA 8 to 20). It is important to keep in mind that the highest possible recommended doses being 0.68 mcg/kg/min and 946 mL/4h, for vasopressors

and IV fluids, respectively, the AI agent (unlike actual clinicians) is limited in recommending very high doses. This may have affected the plots for high severity levels.

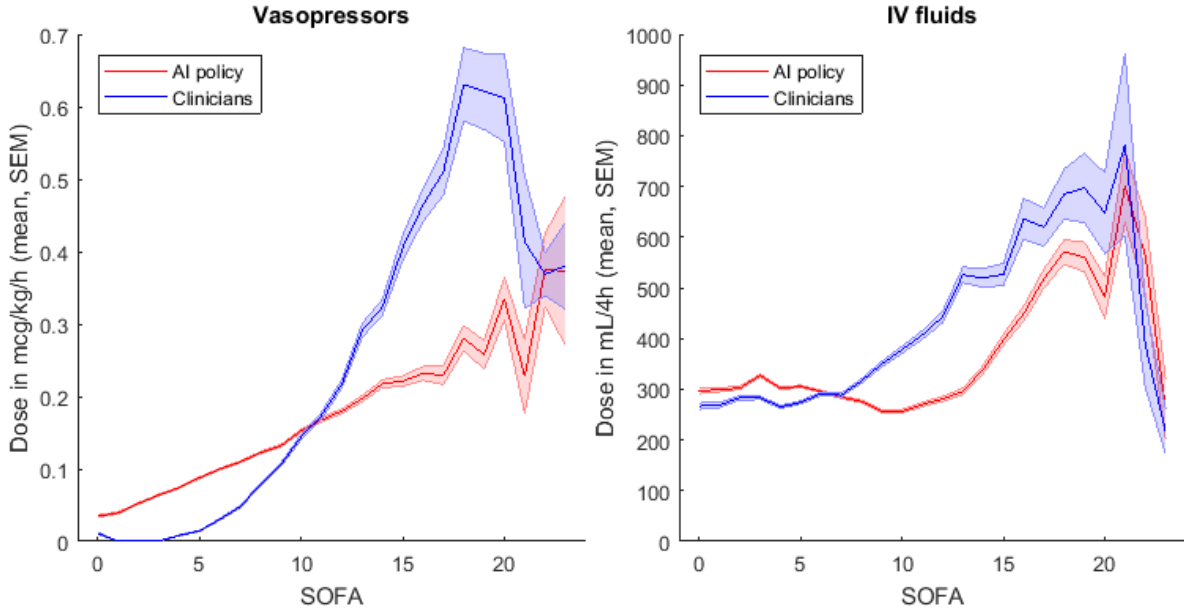


Figure 46: Comparison of actual and recommended dose of drugs as a function of patient severity (approximated by the SOFA score), in the MIMIC-III training set (N = 187,232 samples). Similarly to physicians, the AI agent recommends increasing doses with increasing severity. The AI agent is limited in terms of maximum dose it can recommend, so the rightmost part of the plots (where very high doses are expected) may warrant caution in interpretation. Shaded areas represent the standard error of the mean (SEM). The wide SEMs for high SOFA scores reflect the small number of records in these bins: for example, only 0.1% of the records have a SOFA of 19 or higher.

Behaviour policy estimation

When explaining importance sampling based off-policy evaluation methods, we defined the per-step importance ratio as $\rho_t := \pi_e(a_t|s_t)/\pi_b(a_t|s_t)$. Since it incorporates both the evaluation policy and the behaviour policy, we tested the impact of the variability in the evaluation of the behaviour policy on the WIS estimate using the following method.

In our study, π_b the behaviour policy (clinicians' policy) was computed by observing the actions chosen by clinicians in all the 750 states of the simplified MDP. Formally, we computed the behaviour policy $\pi_b(s, a)$ by counting how many times each action was chosen in all states using the MIMIC-III training

dataset. The counts were stored in an array C . Then, we obtained π_b by dividing C by $\sum_{i=1}^N C(s_i, :)$ with N = number of states, which is the sum of all possible actions taken from state s_i .

The state membership of the training samples is determined by k-means clustering of the training data. Hence, we assessed the impact of different clustering solutions (leading to different behaviour policies) on the WIS estimate. We built 500 separate clustering solutions of the training data and computed the WIS estimate using the same evaluation policy. We show that the 95% lower bound of the WIS estimator exceeds the 95% upper bound of the behaviour policy 332 times out of 500 trials (66.4%) (Figure 47). Among those 500 trials, the median 95% lower bound of the WIS estimator was 82.6 and the median 95% upper bound of the behaviour policy was 58.5. Of course, by design, we select the AI policy maximising the WIS estimator, so the models where the variability in the behaviour policy led to a low WIS estimator were discarded.

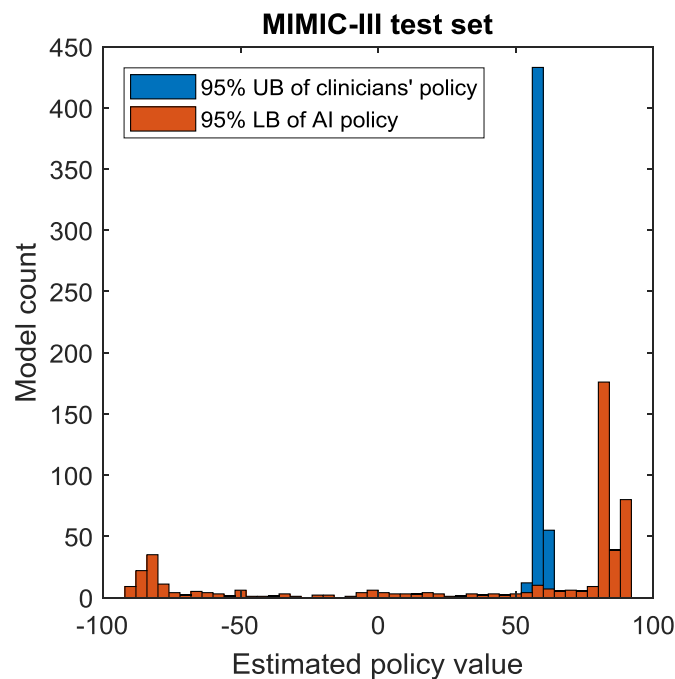


Figure 47: Distribution of the 95% upper bound (UB) of the clinicians' policy (learnt by TD-learning of the Q function) and the 95% lower bound (LB) of the WIS estimator of the AI policy value, in 500 models using different behaviour policies but a unique AI policy, in the MIMIC-III test set. 66.4% of the time, the 95% LB of the AI policy exceeds the 95% UB of the clinicians' policy.

Capture of clinical concepts and diagnoses within the states

As discussed above, clinical concepts such as past medical history and diagnoses are not directly represented as model input features. Using the distribution of ICD codes in the states, we demonstrate that these concepts are encapsulated to some extent within our chosen state definition. Using the 100 most frequent ICD codes in the MIMIC-III training set, we measured how many patients have a given code in all the states, and show the cumulative sum across the states, starting with the states with the highest number of patients with a given code (Figure 48). The black dotted line shows the cumulative sum of patients in the states, with states ordered by descending size. This is the theoretical distribution we should obtain if the ICD codes were randomly assigned to the clusters (the proportion of a given ICD code should be globally equal in all the states). This figure shows that the majority of patients are found in fewer states than if the codes were randomly distributed. For example, 50% of the patients with the ICD code “Coronary atherosclerosis of native coronary artery” are found in only 23 states. This number would be 79 if the code was randomly distributed in the states.

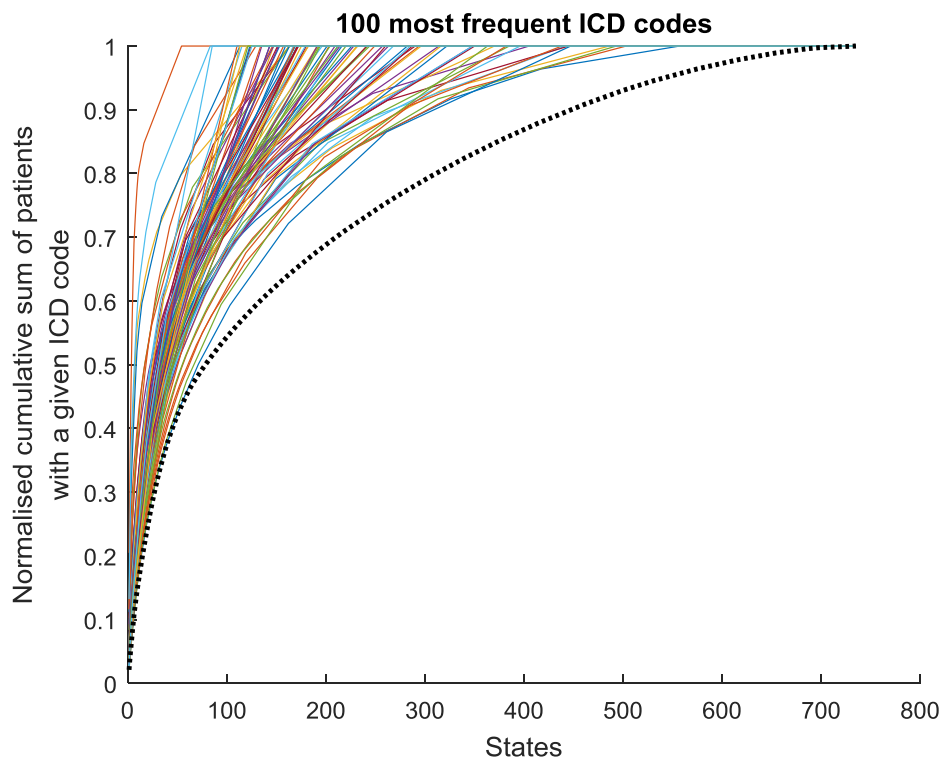


Figure 48: Capture of clinical concepts and past medical history in the model states. We used the 100 most frequently used ICD codes as a surrogate for diagnoses and comorbidities (see text for explanations).

Mortality prediction

Providing an estimate of the expected mortality if the AI policy was used instead of the clinicians' one is an appealing objective, but is probably impossible to achieve without prospective testing. We developed three different methods, which we tested on the MIMIC-III training set. All three approaches yield overoptimistic estimates, with an expected mortality close to zero. From a computational standpoint, this result is not surprising, since the models are – by design – optimising mortality, so this analysis can instead be viewed as a sanity check, which confirms that the algorithm is performing appropriately.

Method 1: Regression model linking policy return and mortality

The idea is to use the regression between the state-action value and patient mortality, since we have demonstrated a monotonous decreasing relationship between both (Figure 37, left). Rather than using visual methods (trying to fit a polynomial to the raw or smoothed data), we can build a predictor (regression model) to learn the relationship between state-action values (from the training data) and outcomes. This can be achieved with any regression model, for example with a random forest (Figure 49) or a logistic regression. We show that the random forest model accurately predicts actual patient mortality in a test set: 22.01% of predicted mortality versus an actual mortality of 22.5%. Then, we can try to predict the expected mortality given the value of the AI policy (learnt by policy iteration). Using the value of the AI policy determined by the model-based approach (median policy value of 82; interquartile range 81.6 – 82.4) or the WIS (median policy value of 84.5; interquartile range 84.3-87.7), the predicted mortality is extremely low, close to 0%.

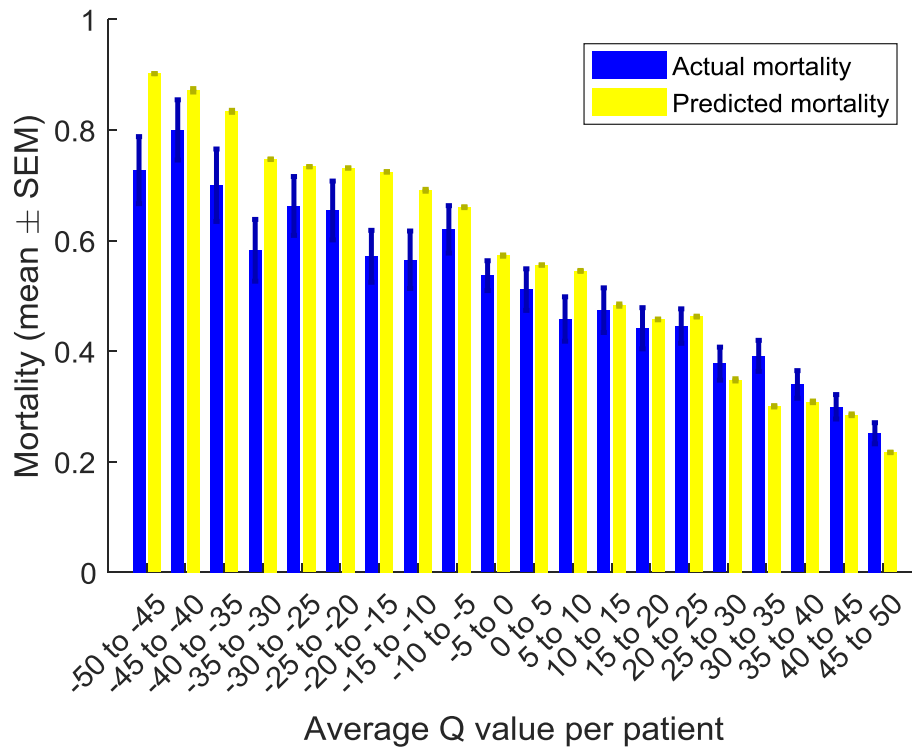


Figure 49: Random forest predicted mortality from the state-action value averaged per patient trajectory. The state-action value function was function learnt by policy evaluation. The plot shows good model calibration.

Method 2: Monte Carlo simulations with actual and optimal actions

Here, we use the model to generate virtual trajectories by following the learned transition matrix and either the clinicians' (Figure 50) or the AI policy (Figure 51) until reaching an absorbing state (death or discharge). We generated 1,000 batches of 2,500 virtual trajectories (for a total of 2.5 million virtual trajectories) to compute the distribution of the estimates. Importantly, this is a model-based approach, which assumes that the MDP has learnt an accurate representation of the reality. As pointed out above, this is not necessarily the case, and indeed apparently perfect models can have poor performance (P. S. Thomas & Brunskill, 2016).

When following the clinicians' policy, the results are extremely accurate (Figure 50): the predicted mortality is 22.47% (SD 0.86%) to be compared to an actual mortality of 22.5% and the average predicted length of trajectories is 14.51 time steps (SD 0.23) whereas it is 14.42 time steps (SD 3.75) in the data.

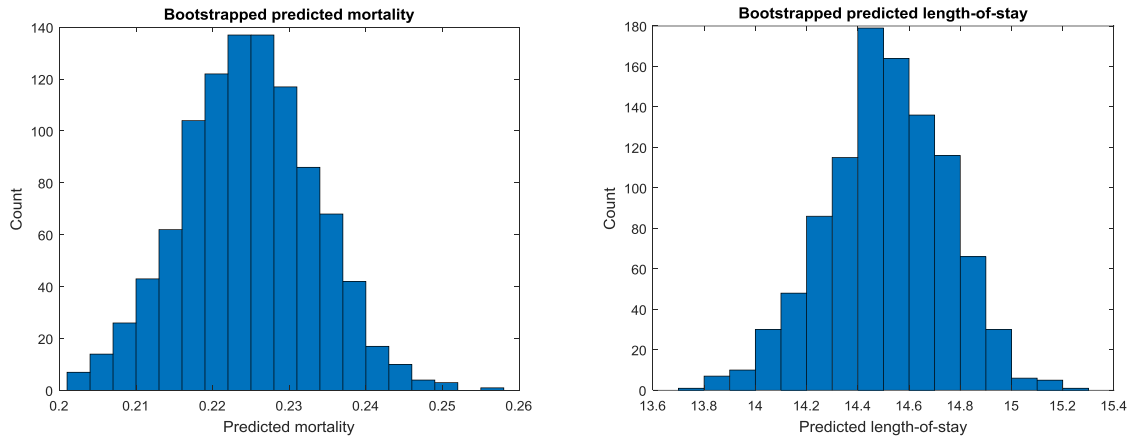


Figure 50: Predicting actual patient mortality and length of trajectories with Monte Carlo simulations.

Then, we repeated the approach when following the AI policy (Figure 51). Here, the decisions taken at each step of the virtual trajectories are the suggested optimal actions. This gives overoptimistic results, of zero mortality and very short trajectories. Predicted mortality is 0.014% (SD 0.037%) and predicted trajectory duration is 4.44 timesteps (SD 0.065).

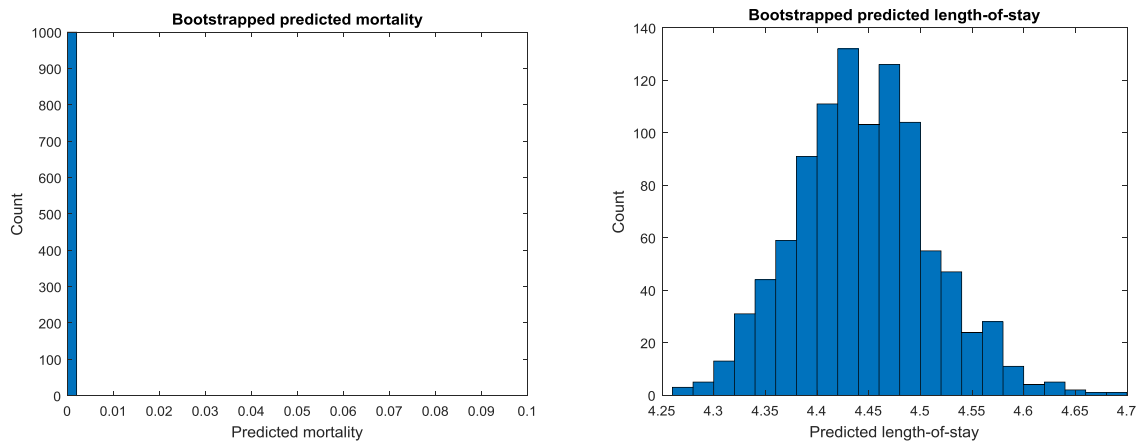


Figure 51: Predicting patient mortality and length of trajectories for the AI policy, with Monte Carlo simulations.

Survival curves can also be plotted (see example in Figure 52, starting in state 333) and show similar results: the expected survival when following only the optimal policy is nearly 100%.

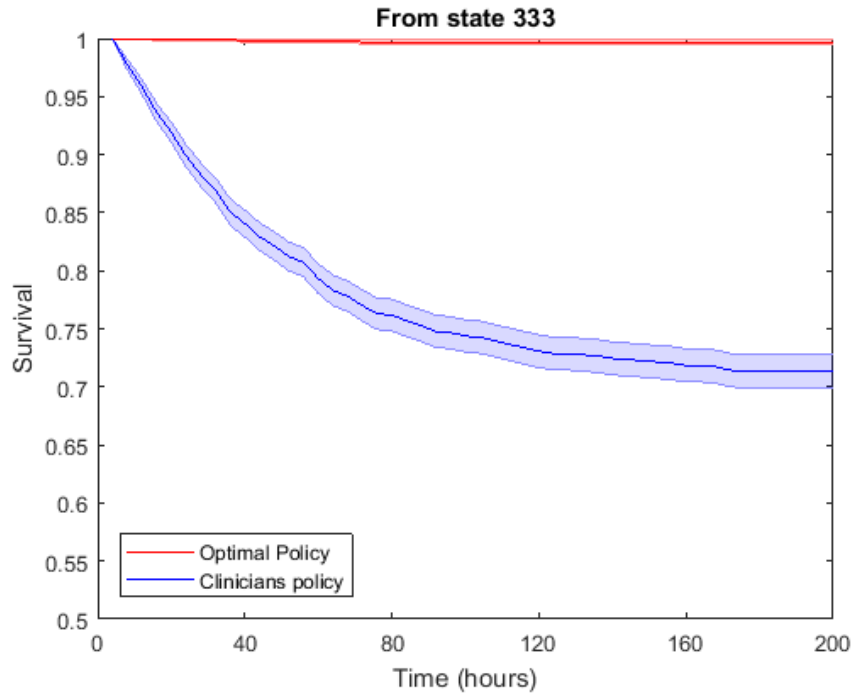


Figure 52: Example of Kaplan-Meier survival curves for the physicians and the optimal policies, built from 2,000 Monte Carlo simulations, starting from the initial state 333. According to this estimate, the optimal policy enables a survival rate of over 99%.

Method 3: Estimation of the policy value from cohort mortality

For a discount factor of $\gamma = 0.99$, an average sequence length of $\bar{T} = 14$ time-steps, a reward of +100 for a patient's survival and -100 for death, an estimate of the value of a policy V^{π_e} as a function of cohort 90-day mortality rate, m , is $V^{\pi_e} \approx \gamma^{\bar{T}}(-100 \cdot m + 100 \cdot (1 - m))$. This value may be made more accurate by taking into account the full distribution of patients' ICU stay lengths (effectively length of trajectories) for both survivors and non-survivors. For our cohort 90-day mortality of 18.9% the expected value of the physicians' policy is 54, which is consistent with the results shown in Table 8 and Figure 35. A policy value of 84.5 corresponds to a mortality rate of 1.5%, which clearly is an overestimation of what could be realistically achieved.

Discussion on model evaluation

A potential bias that could favour the AI agent is that some laboratory values would not have been immediately available to clinicians to inform decision making but were available to the AI agent. The

temporal resolution of our data was down-sampled to 4 hours. The clinical data will, of course, be immediately available and any point of care diagnostics/near-patient testing (e.g. blood gas analysis, glucose) would be available within minutes. Most central laboratory tests (blood counts, biochemistry) would be available within a few hours. Thus, the temporal offset when data becomes available may provide an advantage against clinicians. To estimate the size of this effect we have run a temporal shuffle test, where we shifted “slow” data sources (blood counts, biochemistry) by 4 hours into the future when testing the algorithm performance (as if it were operating on real delayed data), and estimated the value of the AI policy after this manipulation. The result is shown in the next chapter.

Chapter summary

In this chapter, we described the various methods that we developed or implemented to retrospectively evaluate the value of the AI policy. These methods allowed us to select one unique best optimal policy, which we then tested on two independent validation datasets without any further learning.

Chapter 5: Model testing on eRI and CCHIC

After selecting the safest AI policy in the MIMIC-III test set, it was tested with no further learning or tuning on two independent datasets from two different countries: eRI contains 79,073 patients from 128 different ICUs in the US, while CCHIC includes the data from 3,539 patients from 11 separate ICUs in the UK. Both cohorts contain up to 72h of time series of patient data around the time of estimated sepsis onset. This chapter describes the results of these tests.

As explained in chapter 4, the state membership of new test samples was unknown, since these samples were not used during clustering. We determined the state membership of test records according to whichever cluster centroid they fell closest to (using the Euclidean distance), after transformation or normalisation of the features, as appropriate. The knowledge of the state membership is sufficient to identify the optimal suggested action, since the AI policy π^* corresponds to a $\{750 \times 1\}$ vector controlling which action should be taken in every state of the MDP. Bootstrapping was used to compute the distribution of all estimates, and we used 2,000 resamplings (as recommended by Hanna, Thomas and others (Hanna et al., 2016; P. S. Thomas & Brunskill, 2016; P. S. Thomas et al., 2015)), each analysing a subsample of randomly selected patients. For each resampling, we used a sample size of 25,000 patients in eRI and 1,500 in CCHIC.

Model testing on eRI

Value of the clinicians' and AI policies

We estimated the value of the clinicians' policy using TD learning, and the value of the AI policy using 2 approaches: in-place policy iteration (“model-based” evaluation) and importance sampling-based methods (Table 9).

Evaluation method	Clinicians' policy	AI policy
Model-based	54.7 (53.2 – 56.0)	82.0 (81.6 – 82.4)
IS-based	56.9 (54.7 – 58.8)	84.5 (84.3 – 87.7)

Table 9 : Comparison of the value of the clinicians' (evaluated by TD learning) and the AI policies (evaluated by policy iteration – “model-based” or importance sampling – “IS-based”), using bootstrapping with 2,000 resamplings. The numbers represent the median and interquartile range of the policy values.

Model-based policy value estimation

Figure 53 represents the distribution of the value of the clinicians' (evaluated by TD learning) and the AI policies (evaluated by policy iteration), using bootstrapping with 2,000 resamplings. The median Q value of clinicians' and AI policy decisions was 54.7 (interquartile range 53.2 - 56) and 82 (interquartile range 81.6 – 82.4), respectively. According to this estimate, the value of the AI policy was always higher than the clinicians'.

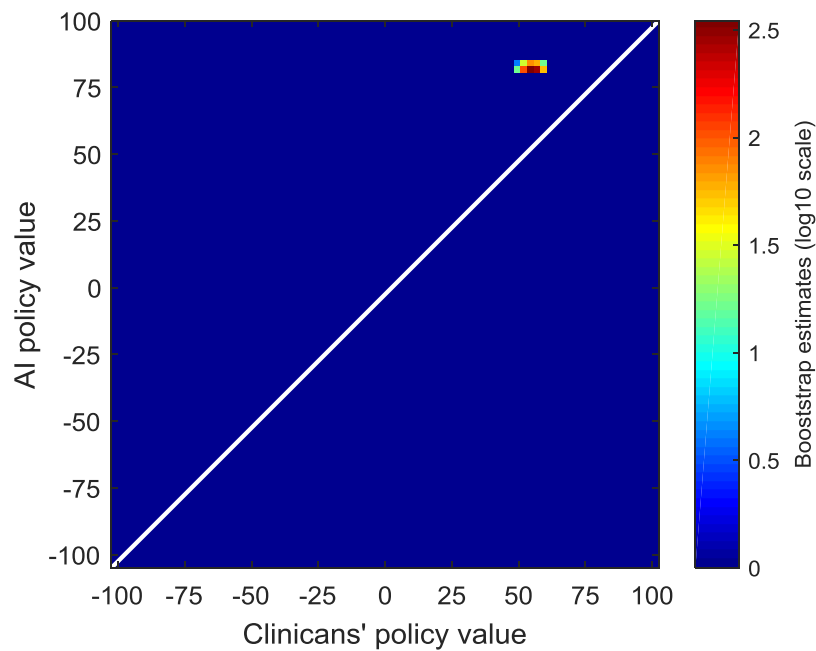


Figure 53: Model-based evaluation of clinicians' and AI policy in eRI. According to this evaluation, the value of the AI policy was consistently higher than the clinicians'.

IS-based policy value estimation

Figure 54 and Figure 55 show the estimated IS and WIS estimators of the AI policy value, respectively. The non-weighted IS estimator led to extreme values of the AI policy. Using bootstrapping with 2,000 resamplings, the median value of the AI policy was estimated at 2.3×10^{12} (interquartile range $2.1 \times 10^5 - 2.3 \times 10^{12}$). In our model, the rewards assigned to the transition to absorbing states bound the range of possible values for any policy, so any value beyond the $[-100; +100]$ range is by definition impossible.

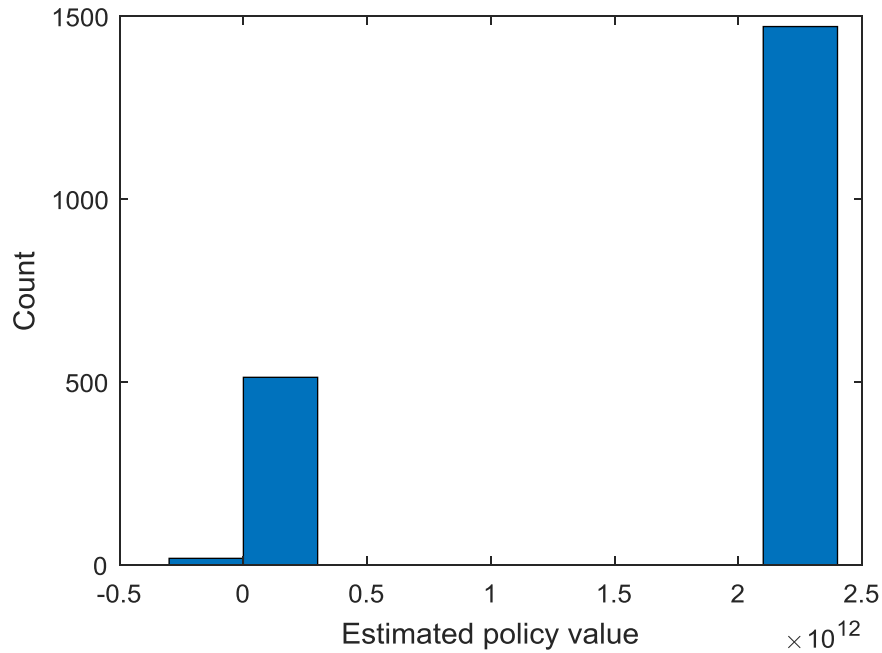


Figure 54: Non-weighted importance sampling-based estimator of the value of the selected final AI policy in eRI, using bootstrapping with 2,000 resamplings. This method led to impossible values, beyond the allowed [-100 ; +100] range.

Therefore, we used the WIS estimator to estimate the value of the AI policy. Figure 55 shows the distribution of the estimated value of the clinicians' policy and the AI policy in the selected final model tested on the eRI cohort. Using bootstrapping with 2,000 resamplings, the median value of clinicians' policy and the AI policy were estimated at 56.9 (interquartile range 54.7 - 58.8) and 84.5 (interquartile range 84.3 - 87.7), respectively. According to this estimator, the value of the AI policy was as good as or better than the clinicians' 96.1% of the time (1,922 times out of 2,000 resamplings).

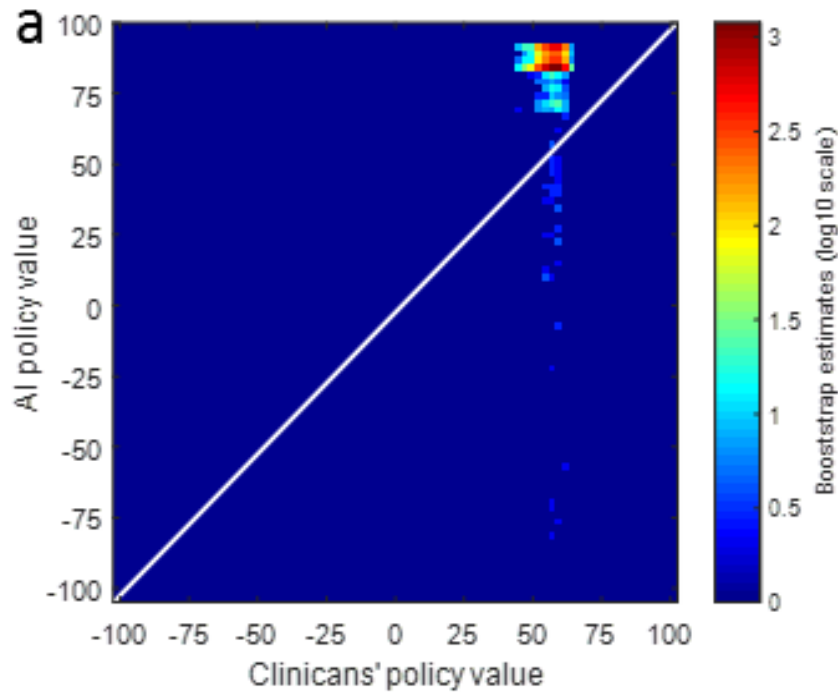


Figure 55: Distribution of the estimated value of the clinicians' and the AI policy in the final selected model, built by bootstrapping with 2,000 resamplings. In this estimation, the value of the AI policy was as good as or better than the clinicians' 96.1% of the time.

Next, we performed a sensitivity analysis to test whether the time at which some laboratory values were available could have favoured the AI agent. Indeed, we wondered whether the temporal offset when data becomes available may provide an advantage against clinicians (this is discussed in detail in the previous chapter). We shifted “slow” data sources (blood counts, biochemistry) by 4 hours into the future, and estimated the value of the AI policy after this manipulation. The performance of the algorithm did not significantly decrease: the estimated value of the AI policy after shuffling the lab values did not significantly differ: 85.1 (IQR 85.1 - 86.0) while it was 84.5 (IQR 84.3-87.7) before.

Qualitative comparison of the clinicians' and AI policies

Figure 56 shows the distribution of treatment doses (the distribution of actions) according to clinicians' and AI policies, in the validation eRI cohort. On average, the AI agent recommended lower doses of intravenous fluids and higher doses of vasopressors, compared to clinicians' actual treatments. While the patients received vasopressors 17% of the time in the eRI cohort, this proportion would have been

35% if following the AI policy. Overall, only 7% of the test records received a low dose of intravenous fluids (corresponding to the first quartile of the distribution of fluid doses, ≤ 50 mL/4h) and vasopressors. This proportion increased to 20% for the AI policy. 33% of the patients received no vasopressor and a high dose of fluid (third or fourth quartile of the fluids' distribution, over 180 mL/4h), this proportion would only have been 23% as per the AI agent.

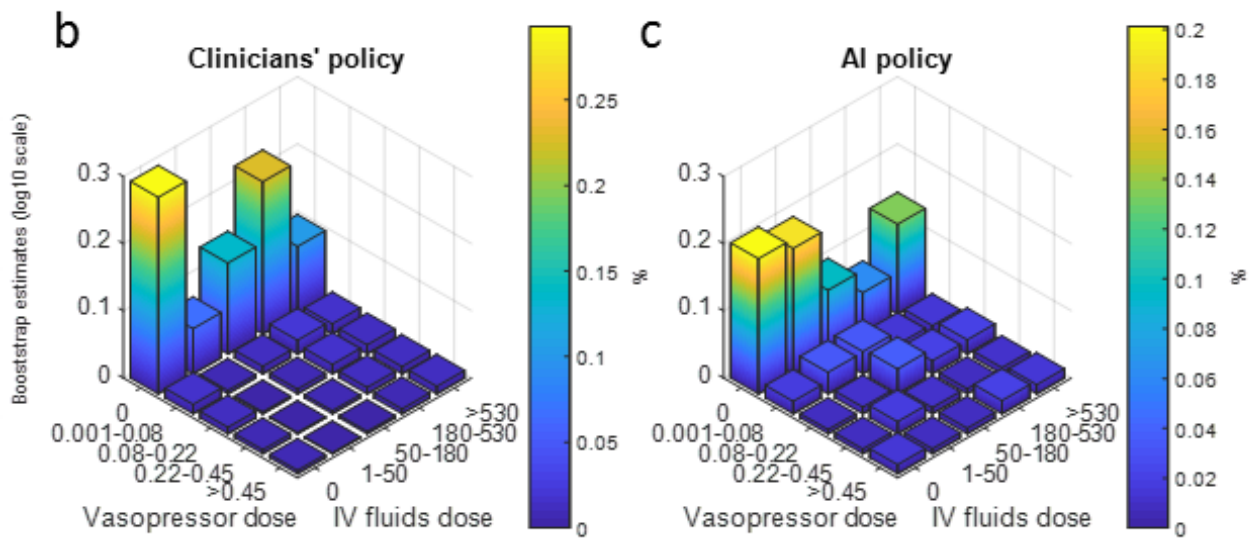


Figure 56: Comparison of clinicians and AI policies in eRI. All actions were aggregated over all time steps for the 5 dose bins of both medications. On average, patients were administered more intravenous fluid (b) and less vasopressor (c) medications than recommended by the AI policy. Vasopressor dose is in mcg/kg/min of norepinephrine-equivalent and intravenous fluids dose is in mL / 4 hours.

For comparison, we provide in Figure 57 the distribution of drug doses for the AI agent without pruning the transition matrix, which restricted the set of possible actions to choose from to frequently observed clinicians' decisions (see the section on the transition matrix, in the chapter about model construction). It is interesting to see that the differences between the clinicians' decisions and the AI suggestions are even more striking: the recommended doses of vasopressors are on average higher, more patients are recommended vasopressors (about 70% versus 30% above), and very few patients are recommended high doses of intravenous fluids (above 180 ml/4h). As discussed in the methods, we chose not to implement this version in the final model since it could potentially be less safe.

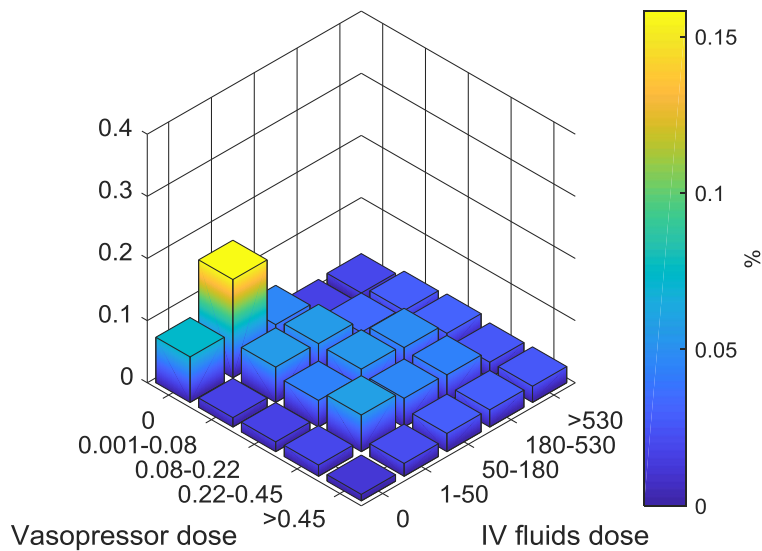


Figure 57: Distribution of drug doses for the AI agent without pruning the transition matrix. When compared to the AI policy after pruning (Figure 56), the doses of vasopressors are on average higher, more patients are recommended vasopressors (about 70% versus 30% above), and very few patients are recommended high doses of intravenous fluids (above 180 ml/4h).

Relationship between the doses received and mortality

We analysed patient mortality when the dose actually administered corresponded to or differed from the dose suggested by the AI policy. Fifty-eight per cent of the time, the patients received a dose of vasopressor very close to the suggested dose, within 0.02 mcg/kg/min or 10% (whichever was smaller). For fluids, patients received the suggested dose approximately 36% of the time, within 10 mL/hour or 10%. These patients, who received doses similar to the doses recommended by the AI policy, had the lowest mortality. When the actual dose given was different from the suggested dose, clinicians gave more or less fluids in similar proportions, and less vasopressors 75% of the time.

Giving more or less than the AI policy of either treatment was associated with increasing mortality rates, in a dose-dependent fashion. Figure 58 demonstrates this association, when the dose gap was averaged at the patient level. The median dose deficit in patients who received too little vasopressors was 0.13 mcg/kg/min (interquartile range 0.04-0.27 mcg/kg/min).

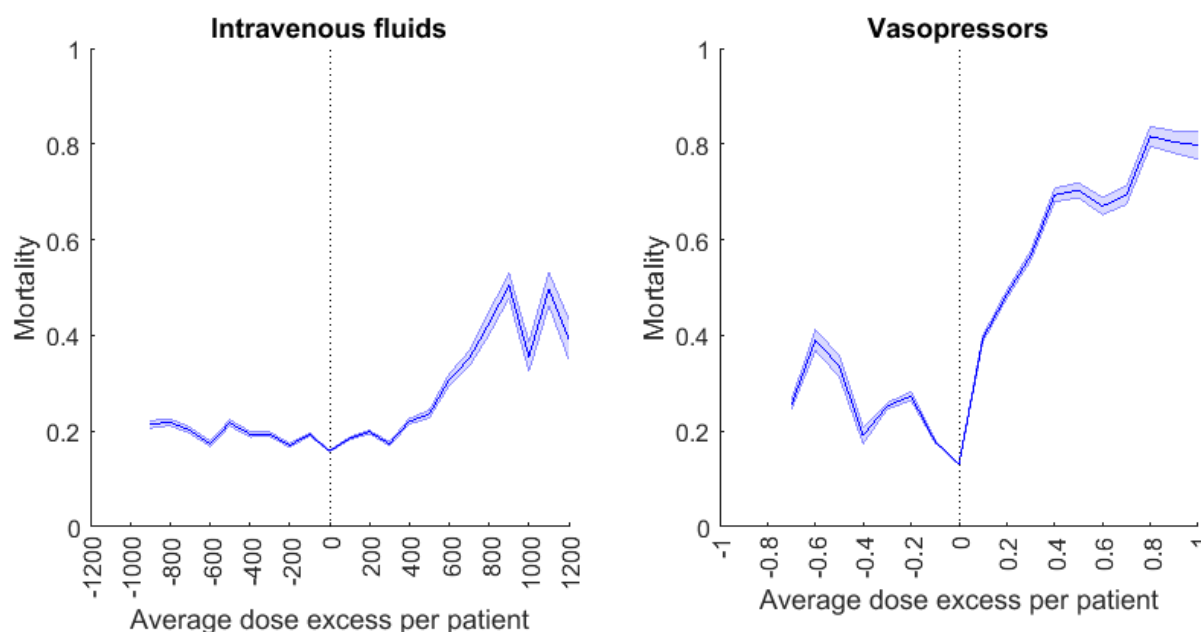


Figure 58: Average dose excess received per patient of both drugs in eRI and corresponding mortality. The dose excess refers to the difference between the given and suggested dose averaged over all time points per patient. The figure was generated by bootstrapping with 2,000 resamplings. In both plots, the smallest dose difference was associated with the best survival rates (vertical dotted line). The further away the dose received was from the suggested dose, the worse the outcome. The shaded area represents the standard error of the mean.

Exploration of individual trajectories

To deepen our understanding of the workings of the algorithm, an important aspect is to scrutinize its behaviour at the level of individual patients and health states. The question we are trying to answer is whether the suggested doses make sense from a clinical standpoint, and whether it is plausible that they may (causally) contribute to the likelihood of transitioning to healthier states. A concern is that optimal actions may sometimes be suggested as a result of “noise” in the data (errors in the state attribution due to poor clustering, itself possibly due to errors in the data, such as outliers), poorly representative subsamples of patients who improve (in any large group, a few patients are likely to experience sudden recovery without being representative) or improvement due to external factors (e.g. new antibiotic added, weaning of sedation or mechanical ventilation...). We explored the trajectories of patients from the training set and present below a few selected noteworthy findings.

We introduce the following presentation to summarise a given state, showing a selection of 19 parameters including total state population, average bloc (time step), average values of a range of patient

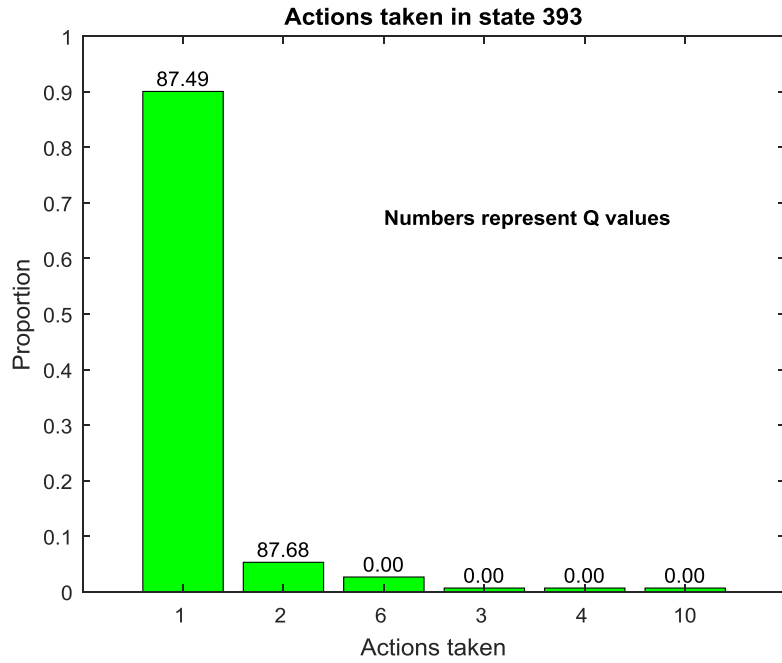
features, average given doses of IV fluids and vasopressors, observed average patient mortality and state value estimated from the policy iteration algorithm. Urine output and fluid intake are expressed as hourly values. “Cumul FB” stands for cumulated fluid balance.

For example, state 393 below appears to be populated by many patients affected by mixed respiratory failure with respiratory acidosis, early on after ICU admission (average bloc of 3.6). Over half of them are on invasive mechanical ventilation, many have impaired renal function (the normal range for creatinine is 0.6 to 1.2 mg/dL) but a preserved urine output. They are not on any fluids and the average dose of vasopressor in this group is very low.

state	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
393	151	3.63	7.82	12.44	80.05	70.23	19.72	7.31	1.37

Cumul									
paCO2	MV	P/F	Creat	UO	FB	Input	vasopr	90d_morta	Value
61.75	0.56	168	2.56	59	-788	0	0.013	0.27	87.50

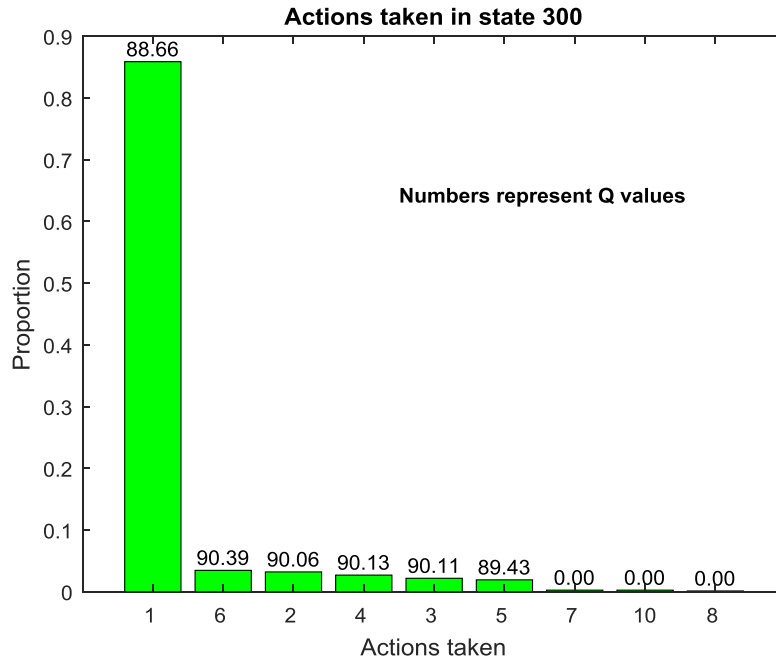
The following histogram presents the distribution of actions taken by clinicians in state 393. 90% of the time, action 1 was taken (no fluid and no vasopressor). Next, action 2 was taken 5% of the time, and action 6 2.6% of the time. The bar also displays the action-values (Q values) for all the actions taken frequently by clinicians. Action 1 has an action value of 87.49 ($Q(393,1) = 87.49$) while $Q(393,2)=87.67$. The other actions don't have a Q value, since we judged that they were not taken often enough to allow inference (we did this by pruning the transition matrix). The optimal action in state 393, as estimated by our algorithm, would be action 2, since it is associated with the highest action-value, among the actions taken frequently by clinicians. One may note that the action value of both actions 1 and 6 are very close.



Let's look at state 300:

state	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
300	779	10.01	3.89	14.67	102.11	74.29	22.02	7.40	1.77

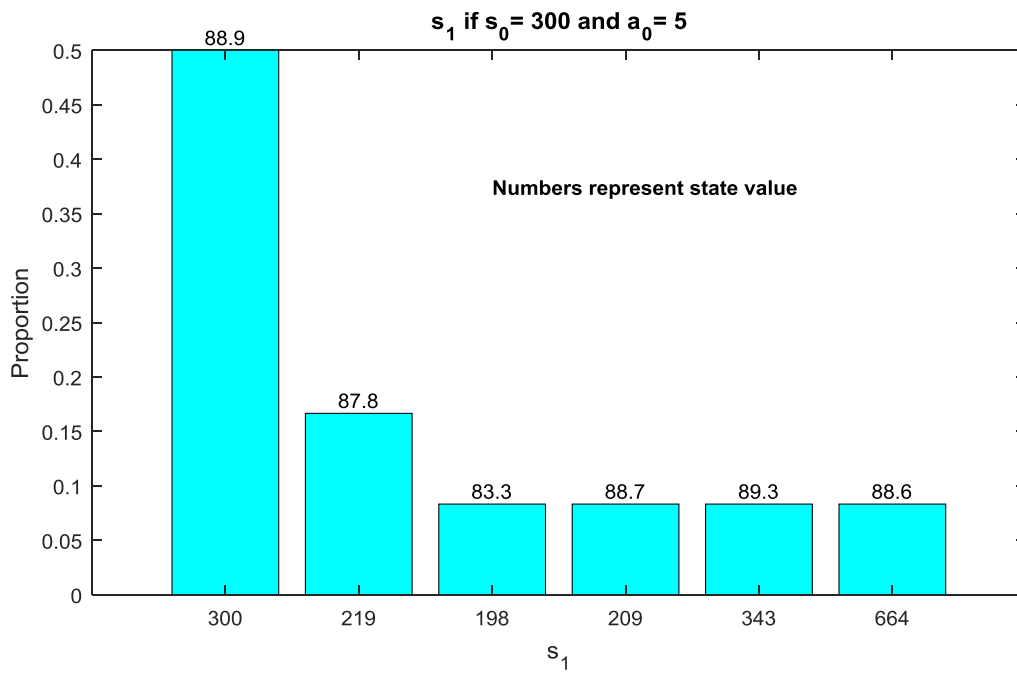
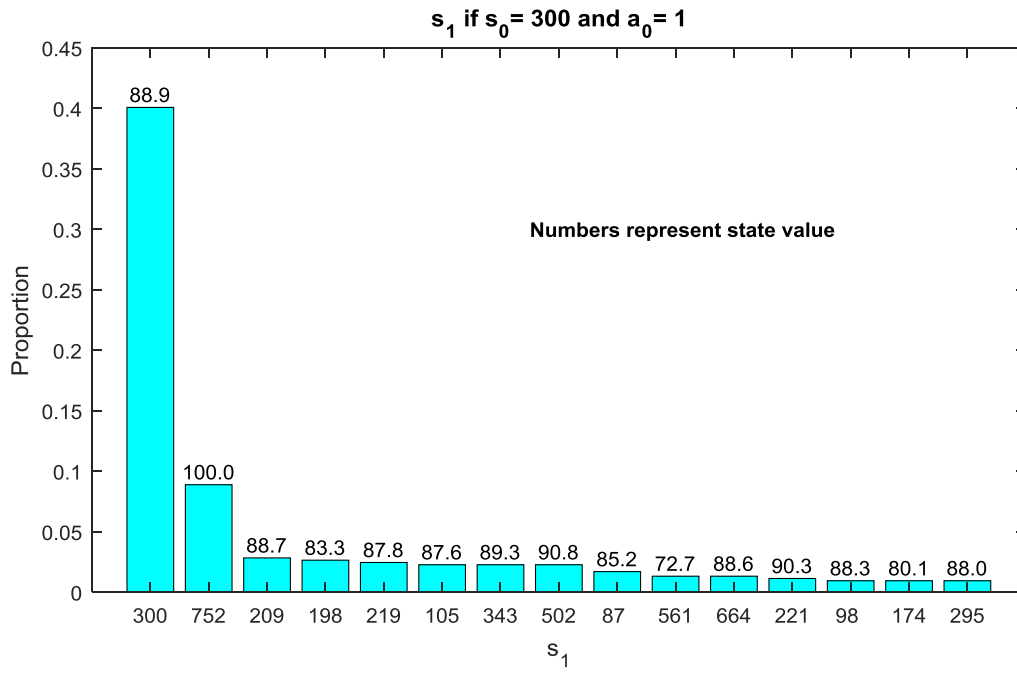
					Cumul				
paCO2	MV	P/F	Creat	UO	FB	Input	vasopr	90d_morta	Value
38.61	0.09	271	1.04	94	2359	0	0.029	0.09	88.86

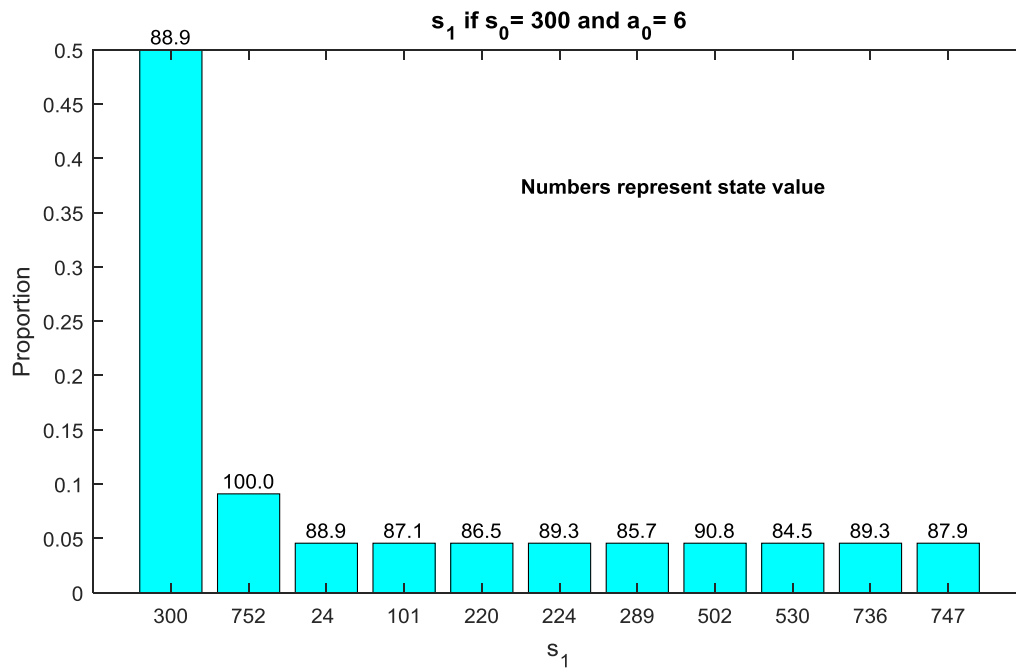


This represents a densely populated state of “late” patients (the average “bloc” is high) with low severity. The rate of mechanical ventilation is low, 86% of the patients received action 1, they are haemodynamically stable, and the average state mortality is very low.

The recommended optimal action is 6 was only taken by 3.5% of the clinicians for patients in this state. We notice that any action from 1 to 6 (meaning any dose of vasopressor and no fluid, or no vasopressor and 30 ml/4h of fluid) has a high Q value (above 88 points), while clearly some of these actions appear clinically odd. Action 5 is probably the most unexpected in this state, but was the option taken for only 1.9% of the patients.

To understand why action 6 is recommended, let’s move ahead one step further, and look at the distribution of states s_{t+1} , when either action 1, 5 or 6 are taken.





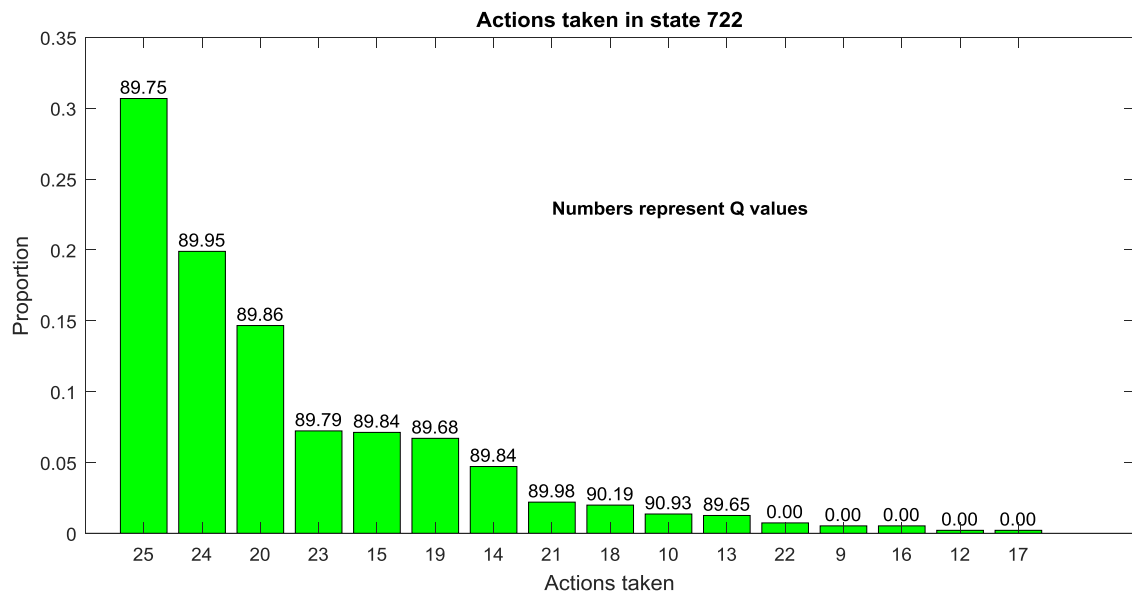
Action 5 (high dose of vasopressor) can't lead directly to state 752 (discharge), which is an important sanity check. Both actions 1 and 6 lead to state 752 in about 9-10% of the time, which helps bringing up their action value. Action 1 can lead to 102 different subsequent states (only the top 15 are shown on the plot), some of which are unfavourable. Overall, the discounted value of taking either action 1 or 6 placed action 6 at the top for the algorithm. The fact that action 1 was taken 85% of the time and action 6 less than 2% of the time was not taken into account, since the algorithm is not designed to have this capability.

Next, we will be analysing state 722.

State 722 is a large state (955 data points) where patients are on a high dose of vasopressors and intravenous fluids, but where the mortality is also very low. They are rather early in the course of the disease (H24 on average), but their fluid balance is already 6 litres positive. They are passing good amounts of urine and are not lactic or overly hypoxic. 80% of them are ventilated.

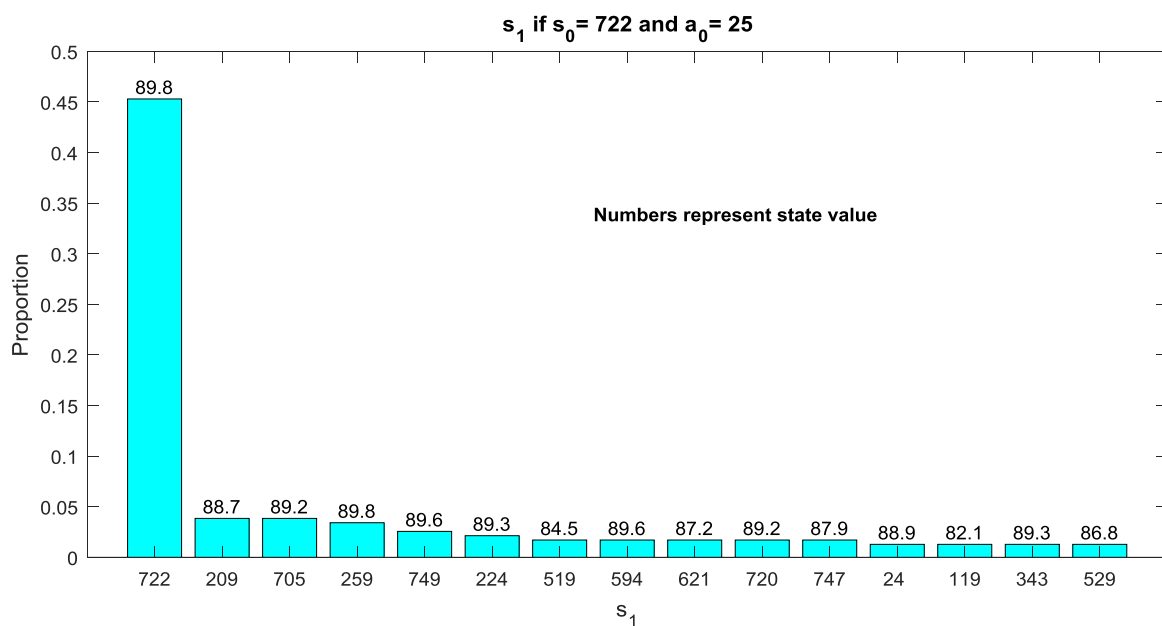
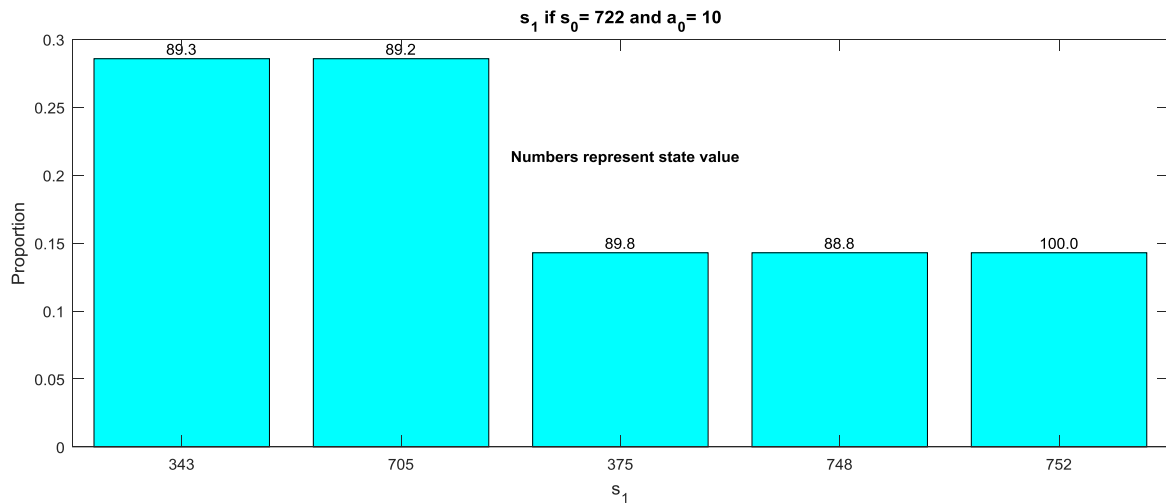
state	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
722	955	6.53	7.35	12.76	81.39	70.08	16.81	7.37	2.01

paCO2	MV	P/F	Creat	UO	Cumul FB	Input	vasopr	90d_mort a	Value
41.07	0.80	297	0.85	102	5948	226	0.581	0.04	89.84



Many different actions were taken in state 722, which is common in large states. The actions taken frequently by clinicians all include high or very high doses of vasopressors, and high doses of intravenous fluids. The suggested action is 10, which is maximum dose of vasopressors and low dose of intravenous fluids. It sounds reasonable from a clinical standpoint and would clearly allow to limit the accumulation of fluids. However, this path was only taken by 1.4% of the patients in state 722 (13 times in the whole dataset).

Let's look at states s_{t+1} , when either action 25 (most frequently used) or 10 (recommended) are taken.



The patients who took action 10 transitioned most often into states 343 or 705. Let's look at the characteristics of these states:

state	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
343	1154	9.15	4.46	14.50	78.57	68.91	17.26	7.38	1.86

Cumul									
paCO2	MV	P/F	Creat	UO	FB	Input	vasopr	90d_morta	Value
42.23	0.11	271	0.89	87	4053	0	0.043	0.02	89.30

state	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
705	1302	7.41	4.38	14.21	83.80	70.64	17.06	7.38	1.93

Cumul									
paCO2	MV	P/F	Creat	UO	FB	Input	vasopr	90d_morta	Value
41.55	0.32	269	0.84	82	4160	5	0.050	0.01	89.20

These two states are similar in the sense that these patients have a low severity of illness. They are large states with a low prevalence of mechanical ventilation, sedation, renal failure, acidosis, and a very low observed death rate. These patients look very different from group 722, from which they transitioned directly. This points at a possible limitation with the current model. These very favourable transitions have occurred in reality in less than 10 patients among several hundreds, but they are chosen by the algorithm since these decisions brought a high reward to the AI agent.

We explored these transitions at the highest possible level of detail and looked at a selection of patients who went through the transition from $s_0=722$, $a_0=10$ into the subsequent states s_1 . We did not identify any obvious, unrealistic issue with those transitions. These patients did not elicit “miraculous” recovery. Rather, they transitioned to health states where they appeared slightly “out of place”. Noticeably, among the 13 patients who took $a_0=10$ from $s_0=722$, the median dose of vasopressor did not change between s_0 and s_1 : it was 0.56 mcg/kg/min.

ICU stay number 45,251, time step number 6:

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
722	6.00	1.00	10.89	88.00	54.19	18.67	7.38	1.58

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
47.00	1.00	262.11	0.90	26	3483	10	0.450	1.00

ICU stay number 45,251, time step number 7:

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
727	7.00	1.00	10.00	77.40	49.47	15.00	7.38	1.60

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
48.00	1.00	215.00	0.90	45	3343	10	0.315	1.00

Patient 76,854, at time step 5:

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
722	5.00	1.00	10.55	84.00	69.27	20.09	7.33	2.46

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
46.64	1.00	314.72	0.60	155	2732	10	0.450	0.00

Patient 76,854, at time step 6:

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
705	6.00	0.00	11.67	88.00	71.17	16.33	7.39	2.45

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
43.00	0.00	372.50	0.60	120	2263	3	0.675	0.00

Those transitions are not erroneous. These patients represent sick patients incoming into states which contain on average relatively healthy patients. We confirmed that these patients did not represent outliers in the incoming state by plotting their position in the cloud of points corresponding to a given state (state 705 in Figure 59), using a dimensionality reduction technique (principal component analysis).

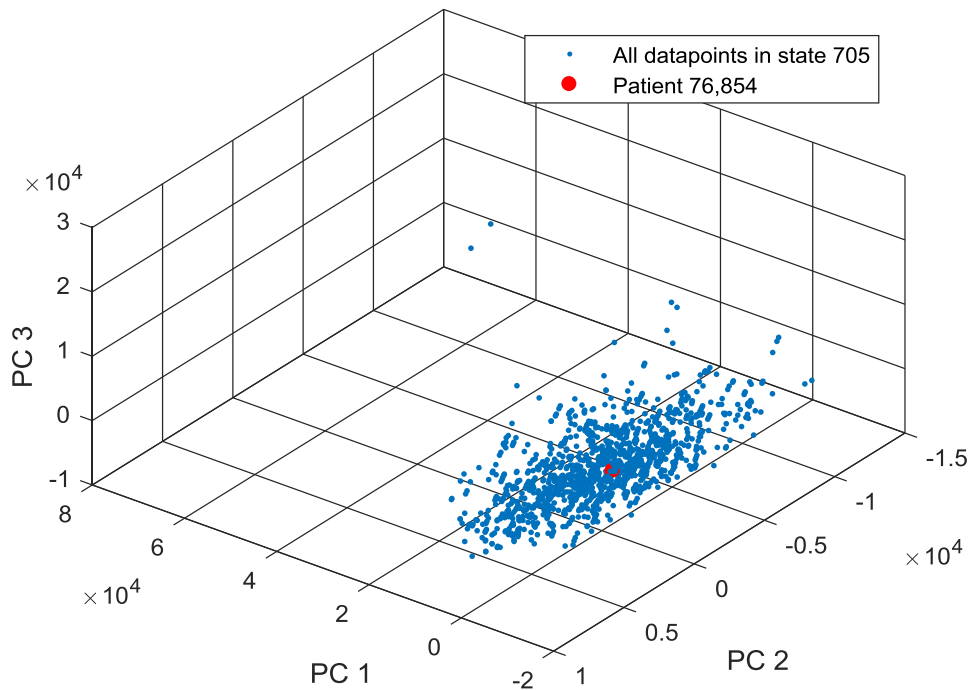


Figure 59: 3-dimensional projection of the first 3 principal components (PC) of patient data in state 705, showing patient 76,854 in red. While patient 76,854 appears clinically sicker than the rest of the state (in particular, he is on a higher dose of vasopressors), the plot confirms that this patient is not an outlier in the incoming state according to our current state definition, since it lies right in the middle of the other data points.

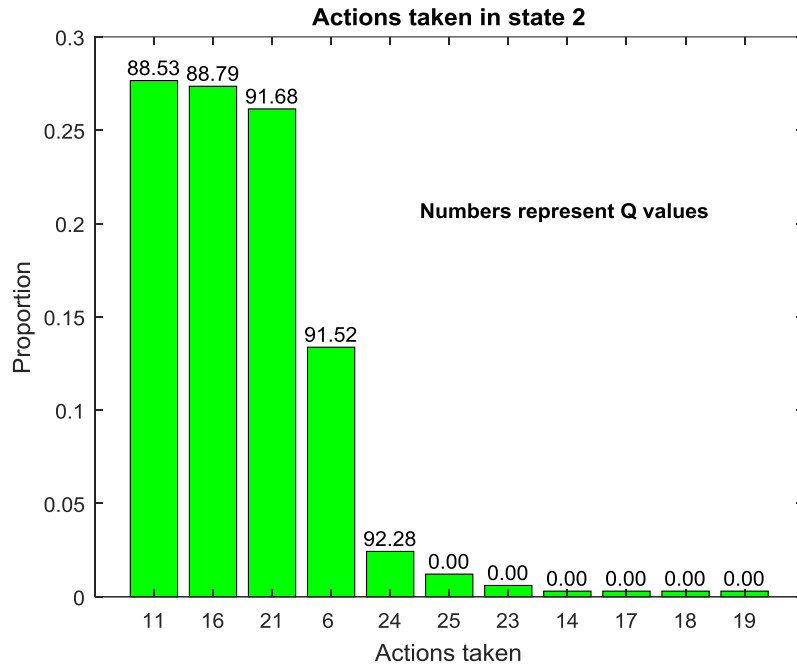
The problem we identified may be that some large states contain an array of patients of various level of severity, including sicker patients. Improving the state definition (clustering mechanism) may be one way to avoid these rare, unexpected transitions. Increasing the number of states beyond 750 may be another way to deal with this potential problem. We discuss this in the final chapter (Chapter 6, “Future directions for model development”).

Let’s look now at state 2:

State	Count	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
2	329	9.94	6.63	11.29	93.29	100.88	20.65	7.44	1.44

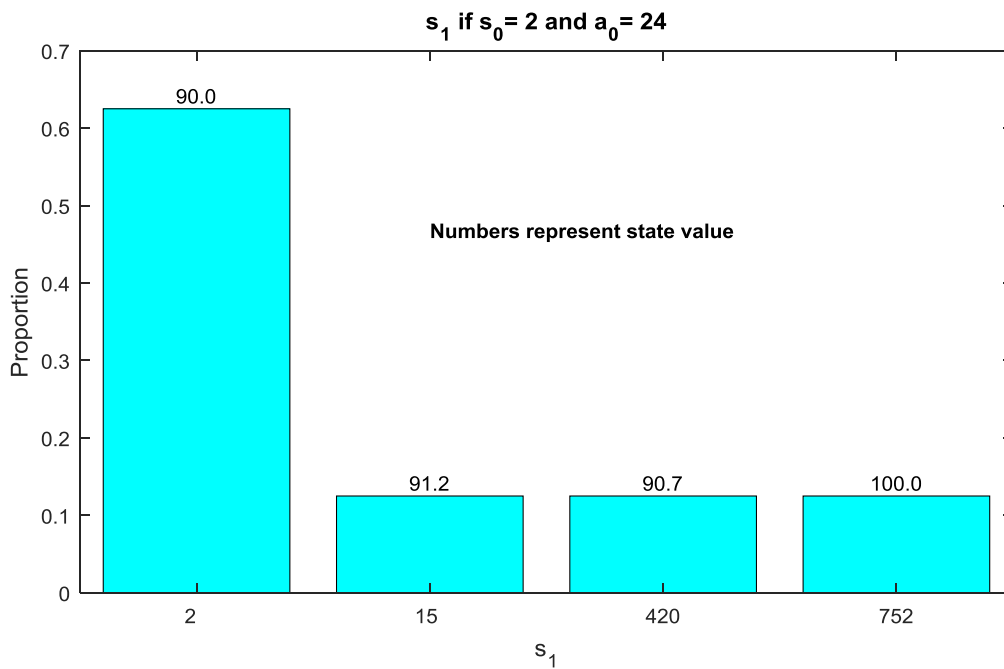
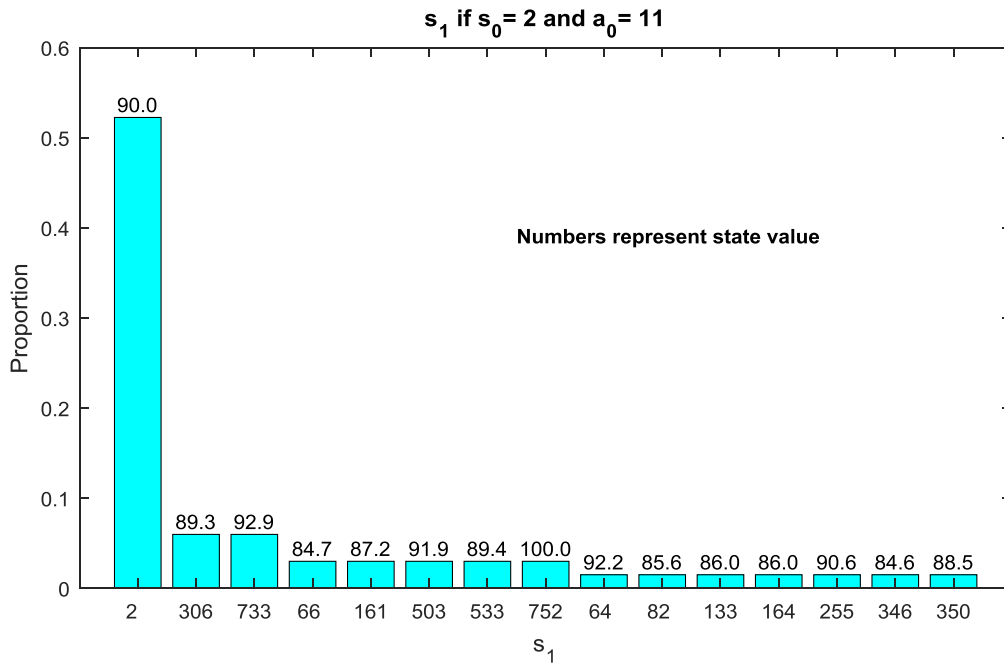
Cumul									
paCO2	MV	P/F	Creat	UO	FB	Input	vasopr	90d_morta	Value
38.20	0.69	232.16	0.96	109	20214	106	0.018	0.14	90.05

In state 2, patients have a high MAP, a very highly positive fluid balance (+20L), are receiving mechanical ventilation 69% of the time, are on a moderate dose of intravenous fluids of 106 ml/h and a low (0.02 mcg/kg/min) dose of vasopressor.



Actions 11, 16 and 21 account for nearly 80% of the decisions made by clinicians. They correspond to zero vasopressor and moderate to high doses of fluids. The recommended action in state 2 is 24 (median vasopressor dose of 0.27 and median intravenous fluids of 946 mL in 4h), which appears unexpected since the average MAP of patients is 100 (which most achieve with no vasopressor) and since the fluid balance of these patients is already extremely positive. Only 2.4% of patients took action 24 in the training data.

Again, let's look at states s_{t+1} , when either action 11 (most frequently used) or 24 (recommended) are taken.



With action 11, three percent of patients were discharged (reached state 752), while they were 12.5% with action 24. Also, no patient who took action 24 died, but one who took action 11 died, which resulted in a penalty of -100 points. Altogether, the algorithm estimated that action 24 was more favourable. The Q values of all the actions (6, 11, 16, 21 and 24) are high, and we can wonder whether the differences in values carry clinical significance or if any choice would have been reasonable in clinical practice. In addition, reaching survival straight after being on high dose of vasopressor and intravenous fluids is unlikely to represent true physiological behaviour. The most likely explanation is

that the data of these patients was censored after reaching the end of the pre-determined data collection period (48h after the estimated sepsis onset).

To confirm this, let's look at the one patient who went through the transition $s_0=2$, $a_0=24$, $s_1=752$. It is patient number 26,936. The last data point is shown below. The patient is still on high dose of vasopressors and fluid input, but the model considers an undiscounted reward of 100 points for action 24 in this instance since the next state is state 752 (survival).

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
2	20.00	0.00	14.67	86.33	116.17	19.00	7.37	0.50

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
44.00	0.00	322.22	1.00	0	30131	205	0.225	0.00

The same phenomenon is seen, for example, with patient 40,264 who is severely ill in his last state (state 602) where action 25 is taken (end of the data recording period) but survives at 90 days.

state	Bloc	SOFA	GCS	HR	MAP	RR	pH	Lactate
602	20.00	2.00	6.25	75.13	118.75	25.00	7.42	12.54

paCO2	MV	P/F	Creat	UO	Cumul_FB	Input	vasopr	90d_morta
42.50	1.00	273.75	0.50	0	24772	313	0.630	0.00

We could present instances of the mirrored phenomenon, for apparently healthy patients at the end of the data collection period but who did not survive to 90 days (which we discussed in Chapter 4).

These investigations led to various important comments about the behaviour of the algorithm and how to improve it:

1. The “dirty gridworld” problem (see section on “Transition matrix” in Chapter 3) was reduced but not eradicated by our approach of pruning the transition matrix. Therefore, one possibility may be to prune the transition matrix further, or to limit the actions available to the AI agent to -say- only the 3 to 5 most frequent actions taken by clinicians. As such, we will limit the opportunity to discover new treatment patterns, but one would argue that we are unlikely to uncover radically new treatment strategies for sepsis.
2. The value assigned to reward and penalty clearly have a strong impact on the behaviour of the agent, which may elicit to skip actions that are associated with good overall survival but occasional (random?) death. A unique death event in a large cohort seem to be able to deter the agent from choosing a path otherwise favourable. Further work in the selection of

optimised reward and penalty values are warranted, and may rely on inverse reinforcement learning to infer the values from directly observing clinicians' behaviour (Littman, 2015).

3. Right censoring of patient data was responsible for spurious transitions from sick states to apparent discharge. A possible solution to this problem would be to discount the reward depending on the time elapsed between the end of the data collection period and the final event (death or discharge). If we apply a similar discount factor of 0.99 for each 4h period, a reward of 100 points would become 74 points after 5 days, 16 points after 30 days and 0.4 points after 90 days, which may preclude the algorithm from converging.

We conducted another analysis, where we computed which proportion of patients took the optimal suggested action, in all the 750 states of the model (Figure 60).

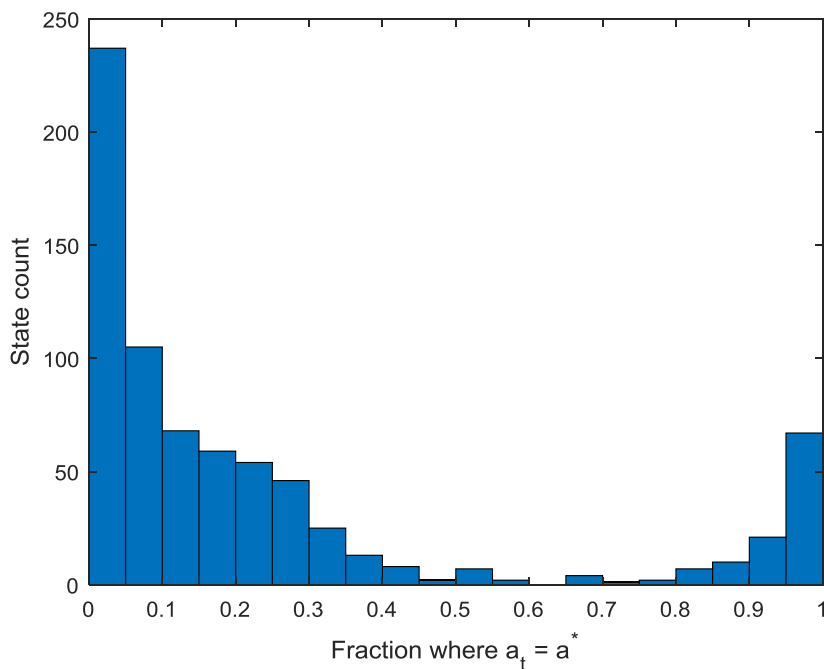


Figure 60: Fraction of patients who took the optimal action, in all the 750 states of the model. In nearly one third of the states, the suggested optimal action was taken by less than 5% of the clinicians in the training dataset. The suggested action was chosen by more than 50% of the clinicians in only 15.6% of the states.

Nearly a third of the time (in 237 states out of 750), the optimal action was taken by 5% or less of the clinicians. In some cases, this may correspond to the phenomenon described above for state 2 or 722,

where paths seldom taken are favoured because they lead to major improvement. We discussed hypotheses for this phenomenon above, and suggested model adjustments to address it. Also, since we envision to develop decision support systems that will be used in real time in the ICU, the percentage of clinicians who took the suggested optimal action could contribute to the level of confidence in the model and be made available to clinicians at the bedside.

Interpretability: relative feature importance

The principle of this analysis is described in the chapter on model evaluation. We estimated the relative feature importance for predicting the administration of the two medications (as a binary feature: drug on or off) using a random forest classification model, and compared results for clinicians to the AI policy. This confirmed that the decisions suggested by the RL algorithm were clinically interpretable and relied primarily on sensible clinical and biological parameters, such as serum lactate levels, mean blood pressure, or urine output.

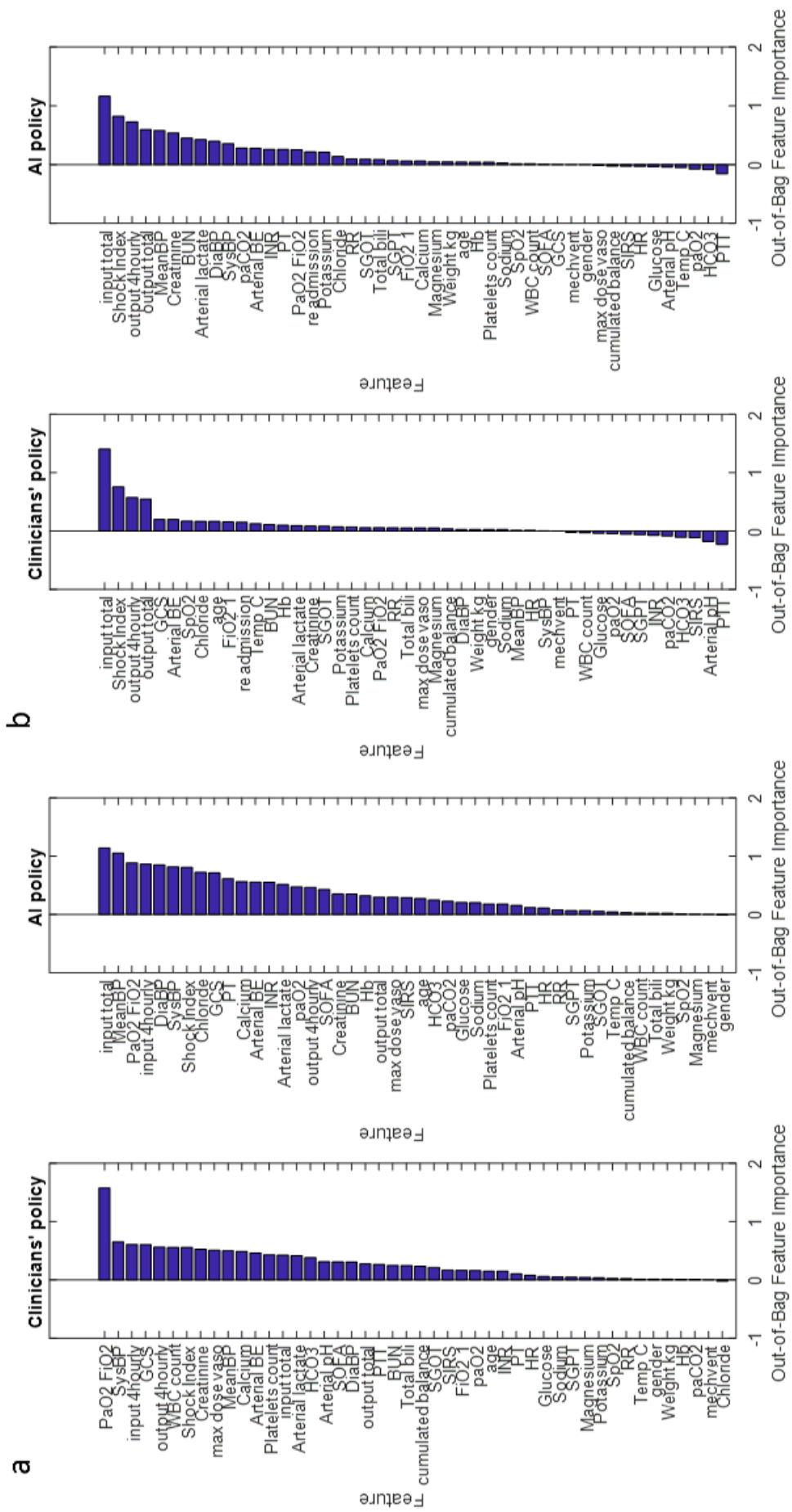


Figure 61. Model interpretability: feature importance underlying the treatment strategies of clinicians and our AI algorithm, in eRI. We built classification random forest models to predict whether both medications were (clinicians' policy) or should have been (AI policy) administered (regardless of the dose), using patient variables as input data. The current dose of vasopressor or intravenous fluid was discarded from the input data in the respective models. Then, the relative importance of each variable was estimated using an out-of-bag technique, where we measured the loss of prediction ability (an increase in the mean squared error on prediction) while we permuted the values of each variable across every observation in the dataset. When permuted, important variables led to large increases in the mean squared error on prediction. Then, we plotted the estimated variable importance averaged over all trees in the random forest ensemble, for intravenous fluids (a) and vasopressors (b), for both the clinicians' and the AI policy. This confirmed that the decisions suggested by the RL algorithm were clinically interpretable and relied primarily on sensible clinical and biological parameters, such as arterial lactate, mean blood pressure, or urine output.

Model testing on CCHIC

Since CCHIC does not contain data about IV fluid administration, it was impossible to fully determine what policy was followed by the clinicians (only the information about vasopressors was available). This limited the extent of the exploratory and validation analyses that we were able to conduct. For example, it was impossible to estimate the value of the clinicians and the AI policy. Also, the cohort definition had to be altered from the strict sepsis-3 criteria, since information about microbiological samples was unavailable (see section on cohort definition). Finally, we identified issues with some of the data fields, in particular with records of vasopressors. It appears that vasopressor infusions were recorded in millilitres per hour for some patients, instead of micrograms per kilogram per minute (the standard unit). From the available data, it was impossible to determine which patients were affected by this problem, so we excluded all patients in which any recorded dose of noradrenaline or adrenaline exceeded 3 (ml/h or mcg/kg/min). However, it is likely than the dose recorded (if in ml/h) was bigger than the dose actually administered for a number of patients that are still in the dataset. As such, the validity of the following analyses is somewhat limited.

Below, we show the analyses that we were able to perform: we report the distribution of vasopressor actions according to clinicians in the CCHIC database (Figure 62), the relationship between dose of vasopressors and mortality (Figure 63) and the relative feature importance for vasopressor administration according to both policies (Figure 64). The methods for generating those analyses have been described in a previous chapter.

Figure 62 compares the distribution of actions of clinicians to the AI policy. We know that the patients in the CCHIC cohort were more severe than the patients in MIMIC-III and eRI, as shown by a higher initial SOFA of 9.4 (standard deviation of 3.5), versus 7.2 (SD 3.2) in MIMIC-III and 6.4 (SD 3.5) in eRI (see Table 3). As such, it is coherent to observe that many more records received vasopressors (about 52% in CCHIC versus 17% in eRI). In this cohort, the AI policy recommends to give vasopressors about 50% of the time, but at a larger dose, on average, than what the clinicians prescribed. Indeed, while only 10% of the records received large doses (dose bins 4 and 5), they would have been 24% according to the AI agent.

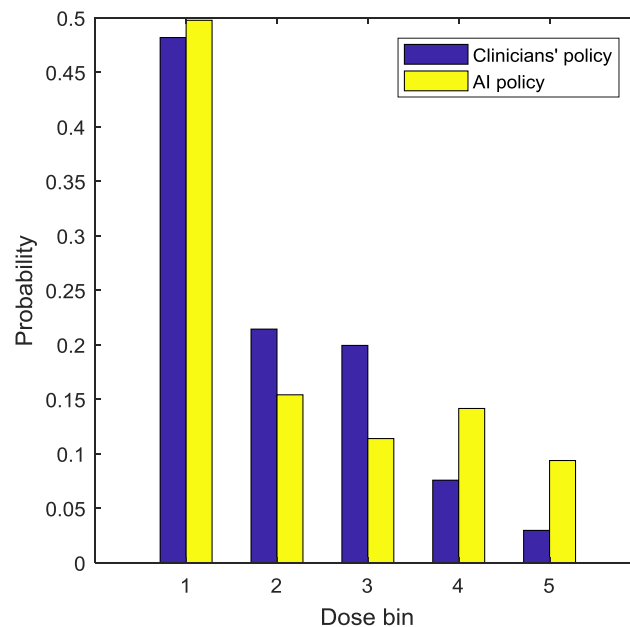


Figure 62: Distribution of the 5 vasopressor actions in CCHIC according to the clinicians and the AI policy. The proportion of time that patients are on vasopressor in CCHIC is around 52%, much higher than MIMIC-III or eRI. While only 10% of the records received large doses (dose bins 4 and 5), they would have been 24% according to the AI agent.

Figure 63 shows the relationship between the dose excess of vasopressors and patient mortality. In this plot, we note a dip for a dose excess of around -0.6 mcg/kg/min. One possible explanation is that the model recommended a high dose of vasopressors (the highest possible suggested dose being 0.68 mcg/kg/min, see Table 5) to patients who did not receive any (hence a “dose excess” of -0.68 mcg/kg/min) and survived. This could represent a possible issue with the approach and hypotheses

include lack of generalisability of a model that was built on a different cohort (US data) and various issues with data quality in CCHIC. The results shown in Figure 64 are satisfactory, and confirm that here again, the model relies primarily on clinically meaningful parameters to suggest a dose of vasopressors.

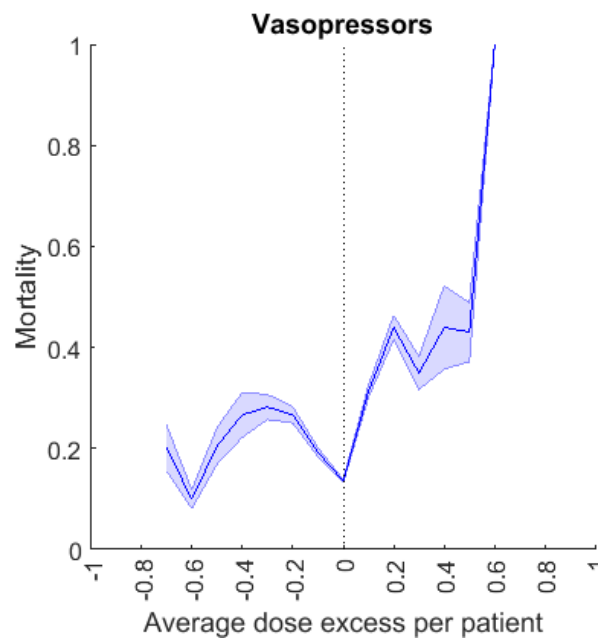


Figure 63: Average dose excess of vasopressors received per patient in CCHIC and corresponding mortality. The dose excess refers to the difference between the given and suggested dose averaged over all time points per patient. The figure was generated by bootstrapping with 2,000 resamplings. The smallest dose difference was associated with the best survival rates (vertical dotted line). The further away the dose received was from the suggested dose, the worse the outcome. The shaded area represents the standard error of the mean. See text for the discussion regarding the dip for a dose excess around -0.6 mcg/kg/min.

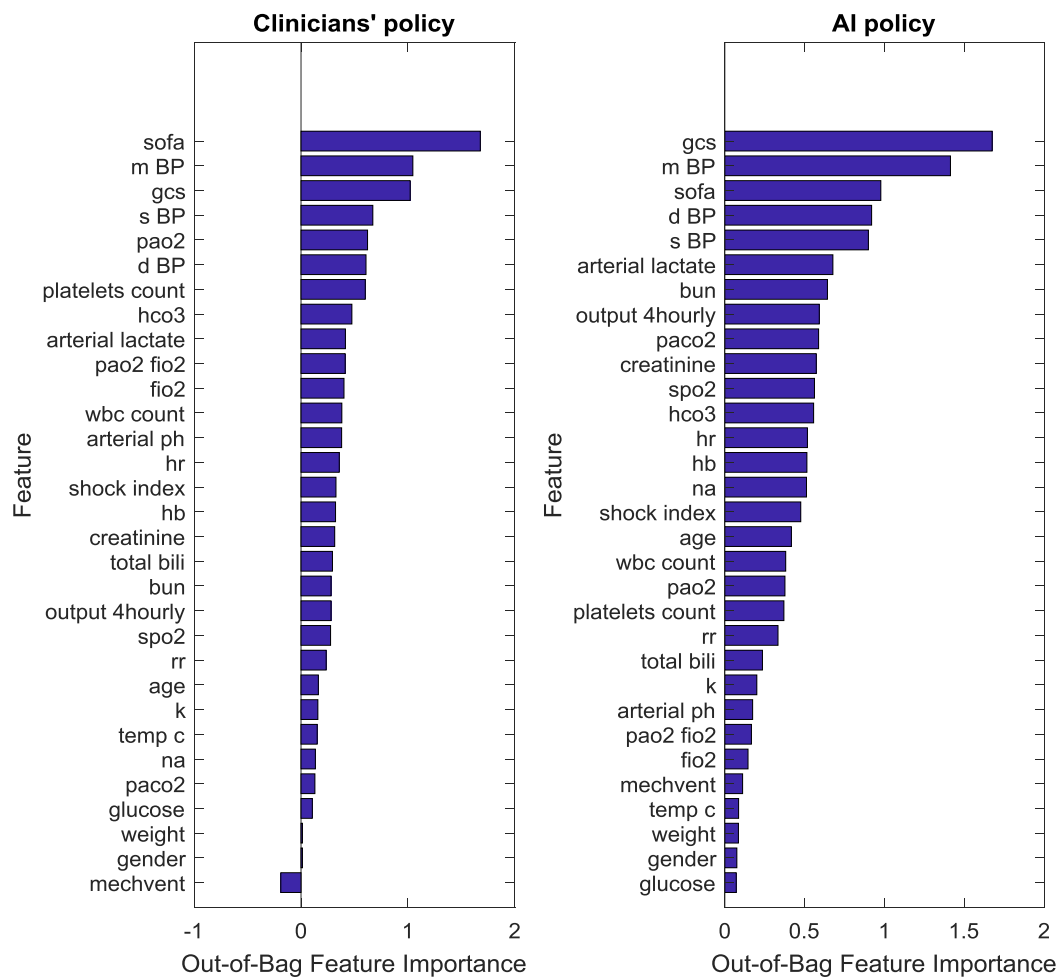


Figure 64: Model interpretability: feature importance underlying the treatment strategies of clinicians and our AI algorithm in CCHIC. This confirmed that the decisions suggested by the AI policy relied mostly on sensible clinical and biological parameters. For both policies, blood pressure and severity of illness are primary indicators for vasopressor administration.

Chapter summary

In this chapter, we showed that the best model learnt from MIMIC-III seems to perform appropriately on two external datasets, with no further learning. The results drawn from the CCHIC dataset need to be interpreted with caution because of potential concerns with data reliability, coming from a database currently in development. The approach proposed for testing new subjects (which relies on determining their state membership in order to identify an optimal dose of drugs) appears promising.

Chapter 6: Discussion

Summary of findings

In this research, we demonstrated how reinforcement learning could be applied to solve a complex medical problem such as fluid and vasopressor management in sepsis, and suggest individualized and clinically interpretable resuscitation strategies. In the retrospective analysis of three large multicentric independent cohorts, the patients that had received the treatments suggested by the learnt policy had the lowest mortality rate.

Previous machine learning approaches achieved above-human performance in pattern recognition tasks for single diagnostic decisions, such as skin cancer or diabetic retinopathy detection (Esteva et al., 2017; Gulshan et al., 2016). Here, we pioneered a different AI approach that goes beyond diagnosis and suggested optimal management strategies. Our AI agent has no explicit knowledge of physiological mechanisms and pharmacology yet it learned from an amount of patient data that exceeds many-fold the life-time experience of human doctors, and discovered an optimal course of treatment for each patient by having analysed myriads of (mostly sub-optimal) clinical decisions. We showed that the computer, which can be considered agnostic, appears to agree with current clinical thinking with regards to the most appropriate way to resuscitate patients with sepsis: a rather conservative approach for intravenous fluids, counterbalanced by widened and earlier indications for vasopressors.

When clinicians' actual decisions varied from the AI optimal policy, this was most commonly to administer too much fluid and too little vasopressor. As detailed in the introductory chapter, it was commonly accepted (at least until recently) that septic patients urgently require large amounts of intravenous fluids during initial resuscitation. However, the harmful effects of administering excessive amounts of fluids and of a sustained positive fluid balance in sepsis are now well documented, and more conservative approaches are called for by many experts (Byrne & Haren, 2017; de Oliveira et al., 2015; Malbrain et al., 2014; P. Marik & Bellomo, 2016; P. E. Marik, 2015). An alternative resuscitation strategy is to use low-dose vasopressor therapy earlier (noradrenaline in the first instance). This is a sensible choice since noradrenaline has many desirable properties that can counter many of the physiologic derangements in sepsis. Interestingly, this is in keeping with our results. While only 7% of the eRI cohort samples received a low dose of intravenous fluids (≤ 50 mL/4h) and vasopressors, this proportion would have been 20% according to the AI agent. While a third of the patients actually received no vasopressor and a high dose of fluid (≥ 180 mL/4h), this proportion would have been less than a quarter (23%) as per the AI policy.

It is interesting to note that patient mortality was only marginally higher when the dose of intravenous fluids given was less than recommended (when the dose excess is negative, see for example the left side of the plot on Figure 58).

This is in keeping with some of the literature suggesting that a delay in administering IV fluids in sepsis may not be associated with an increase in mortality (Seymour et al., 2017). In their large multicentric retrospective analysis, Seymour et al. showed that a rapid completion of a “sepsis bundle” care and rapid administration of antibiotics, but not the rapid completion of the initial bolus of intravenous fluids, were associated with lower risk-adjusted hospital mortality. The effect may persist after initial sepsis resuscitation. An under-dosing by up to one litre of intravenous fluids as recommended by the algorithm has almost no impact on mortality compared to the recommended dose. It can be argued that this does not represent a confounding effect, but that on the contrary, the algorithm learned to capture a debate in the sepsis field. Namely, that it is unclear whether fluid under-dosing/restriction (unlike fluid overdosing/liberal fluid treatment) is associated with worse outcome. For example, the results of the CLASSIC randomized controlled trial in sepsis (Hjortrup et al., 2016) “pointed to benefit with fluid restriction”, where the cumulative fluid volumes received at day 5 differed by 1.2 L (95% CI 0.4 to 2 L) between the treatment strategies. Thus, the “under-dosing” may actually indicate that the AI clinician implicitly learned this fact without being specifically designed to do so, uncovering findings from routinely collected clinical data that were historically obtained from targeted clinical trials. It is unclear why the same behaviour is not seen in the MIMIC-III internal validation test (Figure 38). It may be due to a better performance on internal than external data.

To summarise, our model is in line with the most recent evidence about sepsis management, but importantly allows the treatment decisions to be individualized for each patient. Breaking with previous approaches of sepsis resuscitation, our model does not attempt to correct short-term, arbitrary targets such as mean arterial pressure or fluid responsiveness. Instead, it characterizes relevant signal among a vast amount of medical records matching the clinical state of a new patient, and identifies decisions that are more likely to bring the patient on the path to survival.

Example of a patient trajectory

The algorithm described thus far could be turned into a real-time decision support system, which could provide continuous suggestions of treatment dose of intravenous fluids and vasopressors at the bedside. In order to achieve this, it would be necessary to integrate our model into an electronic health record software, to collect all relevant patient information, matching as many as possible of the MIMIC-III

parameters (see list in Table 1), then to feed them into the model in order to determine the MDP state membership of the patient and its associated recommended dose range of both drugs.

To demonstrate the usability of the algorithm, we selected one patient from the eRI database. The patient features at each time step enabled us to recover the patient state membership, as well as the corresponding suggested action. Figure 65 displays the given and recommended cumulated dose of IV fluids (left) and the dose of vasopressors (right) during the first 48h after the estimated sepsis onset. The shaded area corresponds to the dose range of both drugs suggested by the model (Table 5).

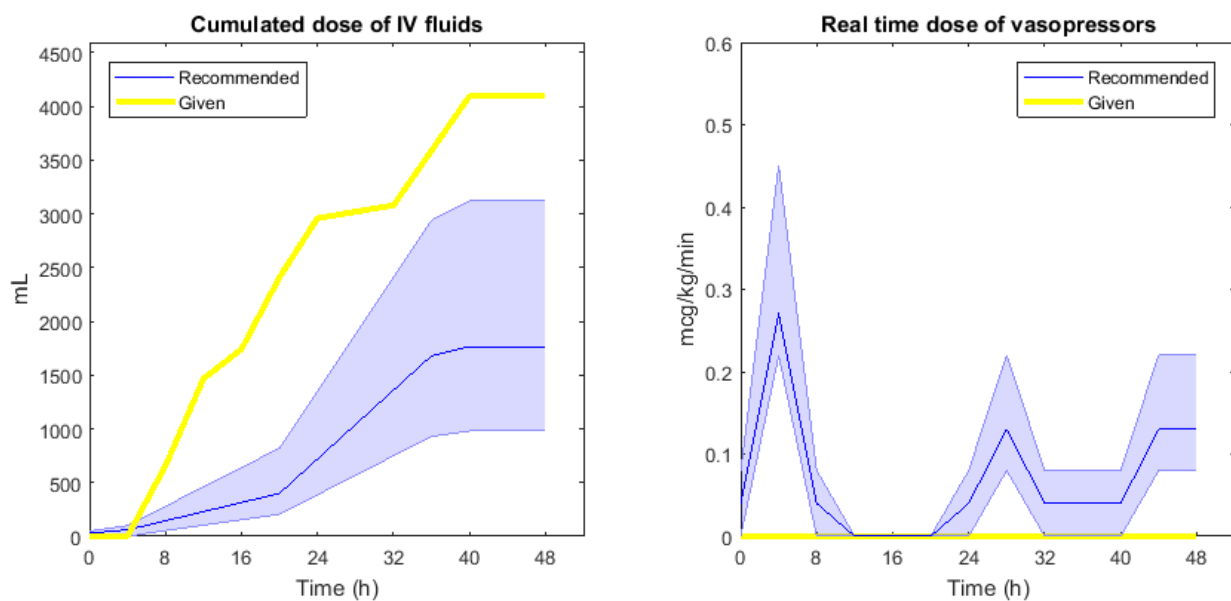


Figure 65: Example of one patient from the eRI database, showing the given and recommended cumulated dose of IV fluids (left) and the dose of vasopressors (right) during the first 48h after estimated sepsis onset. At each time step, the model suggests a dose range of both drugs, which is represented by the blue shading.

Strengths of the approach

To the best of our knowledge, this is the first instance of the successful application of reinforcement learning to address the question of sepsis resuscitation, which is a global healthcare challenge and a major source of healthcare expenditure. Our research surpassed the current state of the art in both health data analytics (which deals with real-world problems but is often limited to observational studies or predictive models that do not provide direct assistance with decision making) and fundamental machine

learning research (which is often limited to artificial tasks) (D'Agostino RB, 2007; M. Ghassemi et al., 2015).

The size, multicentric and international nature of the datasets we used in the research is a significant improvement over previous attempts. Our model represents a compelling example of how machine learning, when deployed by data scientists with domain expertise, could finally enable the development of a truly personalised medicine and improve patient outcomes (Beam & Kohane, 2016; J.-L. Vincent, 2016).

We showed that the model is robust to variability in the features' availability and missingness in the data, and also quantified the relative importance of each feature for determining the state membership. The modelling choices we made rely on relatively few assumptions (such as the number of states, the balance between reward and penalty, the value of the discount factor γ , etc.).

The RL approach that we developed is agnostic to the clinical question and the type of data used and could in principle be applied to any data-rich clinical environment and any medical intervention. Examples in the ICU include mechanical ventilation, sedation, antibiotics administration, among many others.

Limitations of the approach

The model built and published represent an early attempt at applying reinforcement learning to address a highly complex medical question. While the work was very innovative in the way we developed and applied machine learning methods, many aspects of the model could be improved by us or others. There are undoubtedly limitations to our approach, some of which have been discussed above in the separate chapters.

Representing the complexity of a patient in intensive care using only numerical values and a very limited set of interventions is clearly an oversimplification of the reality. Many events and interventions that can completely change the trajectory of a patient were not captured by the model (e.g. sedation, mechanical ventilation, high levels of PEEP, a surgery, a massive blood loss, the prescription of vasodilators, a cardiac arrest, the initiation of haemodialysis...). Even if the size of the array of variables represents a major improvement compared to previous attempts, clinical signs, diagnoses and past medical history are poorly accounted for, mostly because they are not numerical, but also because the reliability of their recording in the EHR was sometimes doubtful. The validity of diagnostic coding in the ICU has been questioned (Misset et al., 2008). To some extent, the use of the Elixhauser comorbidity

score enabled capturing patients' premonitory status, and we demonstrated above that clinical concepts were partially encapsulated within the state definition that we used. We discuss the addition of further patient features below.

When exploring individual patient trajectories, some results pointed to the possibility of spurious transitions and remaining confounding between health state and treatment indication. We propose below (in the section "Future directions for model development") some potential adjustments that could improve the problem.

Some categories of patients require dedicated management strategies, including blood pressure targets, fluid balance adjustments etc. Examples include patients with traumatic brain injury, cardiogenic shock, or chronic renal failure (some of which have no residual diuresis). One could argue that since these patient populations are treated differently than the other patients, they corrupt the learning of optimal policies, and as such should be excluded, or that targeted policies should be developed.

We can discuss the choice of the vasopressor selection, since the drugs included all have different properties but were aggregated into a unique noradrenaline-equivalent. Simplifications had to be made to make the problem tractable, and we used previously published dose correspondence (Brown et al., 2013). Also, noradrenaline was very prominent and represented over 80% of the vasopressor records.

We could have expanded the action space, in particular to include dobutamine for sepsis induced cardiomyopathy. Indeed, between 20 and 60% of patients with sepsis present an associated cardiac dysfunction, which may require specific targeted management with drugs such as dobutamine or levosimendan (Levy et al., 2018; Sato & Nasu, 2015). A rapid exploration of the MIMIC-III data found that the amount of records of dobutamine in mimic is only about 10% of the amount of noradrenaline records, a large proportion of which can be expected to be given to postoperative elective cardiac patients. It appears unlikely that it would be possible to expand the action space to include dobutamine.

The strict requirements we fixed in terms of data quality for model building and testing mean that some sites and patients had to be excluded. For example, about two thirds of the eRI patients were discarded because their data was not of suitable quality. The final cohorts were still relatively large, totalling over 720 continuous years of cumulated data.

Due to differences between the datasets, slightly different implementations of the sepsis criteria were used. Instead of 90-day mortality, hospital mortality was used in eRI and CCHIC.

Rather than using MIMIC-III for model building, we could have used eRI, since it is the largest dataset and offers the most practice variability. The reasons were numerous. First, only MIMIC-III had 90-day

mortality available, which was the outcome on which we wanted to focus. Next, a practical reason is that we had access to MIMIC-III earlier than the eRI. Finally, another rationale is that we can argue that it is more valuable to test a model on a very large dataset collected from over 120 centres than validating it in a single centre in Boston (MIMIC-III).

Future directions for model development

A number of directions could be explored to improve the current work or pivot towards more complex RL models.

We could try to restrict the feature set only to parameters selected by expert opinion (providing “expert priors”), or on the contrary incorporate new features into the state definition, and see whether this could improve the model (for example with regards to off-policy value of the AI policy).

Restricting the feature set could be appealing. Indeed, we have shown that the state definition could be improved, since we have identified patients in large states that did appear slightly different from the rest of the patients in the cluster, from a clinical standpoint (see section on “Exploration of individual trajectories” in Chapter 5). Restricting the set of features to the ones that appear the most relevant may increase the relative weight of each remaining feature. We could choose to use a weighted clustering algorithm, and put more emphasis on the features that we judge important (e.g. MAP, urine output, current dose of vasopressor, etc). Increasing the number of states may also help with this issue of apparent heterogeneity within states.

If, on the contrary, we wanted to enrich the feature set with new data, an obvious approach would be to add the unstructured information found in the clinical notes, nursing notes, radiology reports etc. to help in defining the patient state. As discussed in the section on “data requirements” in Chapter 2, using textual information for patient state definition, while it appears appealing, is not without drawbacks. First of all, there is no established method to achieve this. It is a very difficult task, and an area of active research (M. Ghassemi et al., 2014; Miotto et al., 2016; Rajkomar et al., 2018). Various methods to generate “latent feature representations”, including deep learning auto-encoders and Latent Dirichlet Allocation (LDA) have been proposed (M. Ghassemi et al., 2014; Miotto et al., 2016; Rajkomar et al., 2018; Suk, Lee, Shen, & The Alzheimer’s Disease Neuroimaging Initiative, 2015). Rajkomar et al. demonstrated that it was possible to use unstructured patient notes, for example to predict hospital mortality (they achieved an AUROC of 0.93–0.94). While it could be valuable to incorporate such information as “latent patient representation”, this poses two immediate problems: 1) the need to have good quality (informative) clinical notes available, which is often not true shortly after admission; and

2) the low sampling frequency of this type of information (at best a note is available a few times per day). Another simpler approach would be to convert some of the textual information (e.g. past medical history or diagnosis) into binary or categorical variables. The problem here is that we would end up generating a vast number of these new variables, while actually the gain in information would be very modest since significant simplifications would be required (e.g. summarising all the existing chronic cardiac conditions into one binary feature called “cardiac history”).

When exploring individual patient trajectories, we identified several potential ways to improve the behaviour of the AI algorithm in the current model. To summarise:

1. The “dirty gridworld” problem (see section on “Transition matrix” in Chapter 3) was reduced but not eradicated by our approach of pruning the transition matrix. It may be necessary to prune further the transition matrix, and to limit the actions available to the AI agent to -say- only the 3 to 5 most frequent actions taken by clinicians.
2. A unique death event in a large cohort seem to be able to deter the agent from choosing a path otherwise favourable. Further work in the selection of optimised reward and penalty values are warranted, and may rely on inverse reinforcement learning to infer the values from directly observing clinicians’ behaviour (Littman, 2015).
3. Right censoring of patient data was responsible for spurious transitions from sick states to apparent discharge, which perturbed learning good policies. We should discount the reward depending on the time elapsed between the end of the data collection period and the final event (death or discharge). Excessive discounting may preclude the algorithm from converging.

Other modelling choices could be explored. Instead of a simplified discrete state and action MDP, we could attempt to formalise the problem as a POMDP or a continuous state/action MDP. We discussed the challenges of solving POMDPs, particularly pertaining to their intractability with brute-force methods due to the infinite number of possible belief states. New methods have been developed since we laid the foundation of the initial MDP model. For example, Li was able to develop an actor-critic algorithm including a heuristic search tree to approximate the belief state, which rendered the problem tractable (Li et al., 2018).

A continuous state space implies an infinite number of possible states, so a common approach is to use deep RL techniques to learn Q values or the optimal policy directly from input features (Mnih et al., 2015). In this approach, the patient data would be fed directly into a convolutional neural network in order to rank the actions possible to perform in that state, without the need to discretize the states (Figure

66). Said otherwise, such a model attempts to predict an optimal action given a vector of patient parameters, without explicitly defining a state membership.

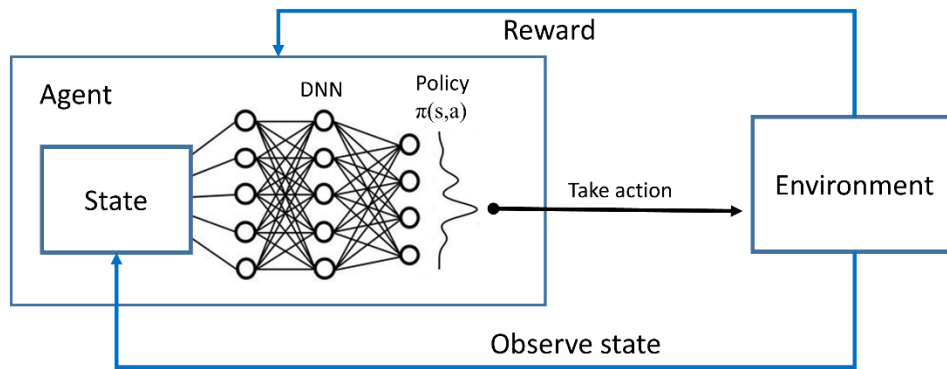


Figure 66: Concept underlying deep reinforcement learning. Here, the optimal policy is learnt directly from input features by a deep convolutional neural network, without discretization. DNN: deep neural network.

Another possible direction could be to assess mixture-of-experts models, which blend several policies to better tailor an optimal strategy (see illustration Figure 67). This has proven to work well in other medical problems such as selection of HIV treatment (Parbhoo et al., 2017). The authors suggested that the reason for the improved performance is that it may be difficult for a single model to perform well on all types of patients, due to their heterogeneity. The mixture-of-experts framework that was proposed combined a kernel-based expert (a neighbour-based policy maximising the survival rate) and RL-based (in fact, deep RL) expert, and uses a gating strategy to switch action between policies suggested by both experts.

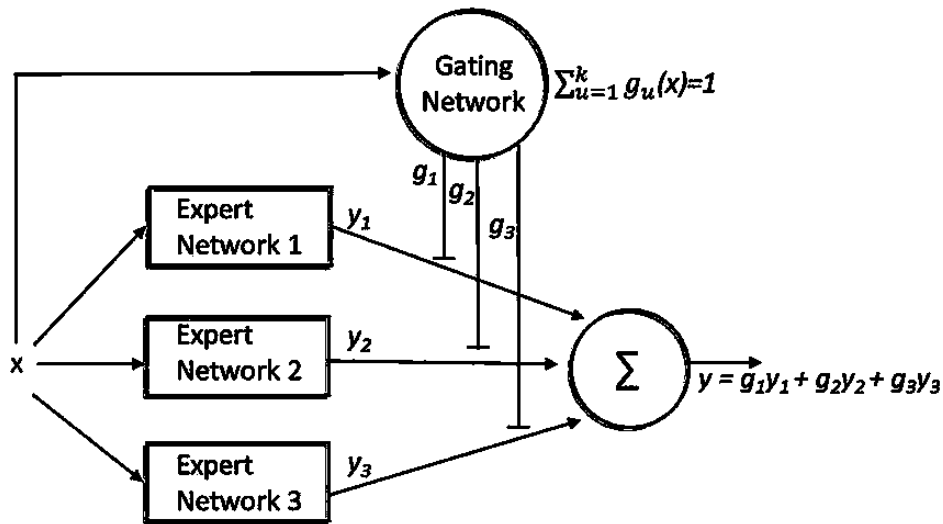


Figure 67: Example of a mixture-of-experts architecture, where the expert networks compete to learn a task while the gating network mediates the competition. Reproduced from (Bock & Fine, 2014).

Towards clinical use

The work presented here is clearly not ready for clinical deployment. We welcome and encourage scientific scrutiny of the model, which will require further refinement and testing before cautious prospective testing for safety and effectiveness before considering clinical use. While physicians will always need to make subjective clinical judgments about treatment strategies, we believe that computational models can provide additional insight about optimal choices, avoiding targeting short-term resuscitation goals and instead following optimal trajectories towards longer-term survival (Chen & Asch, 2017; A. E. W. Johnson et al., 2016; J.-L. Vincent, 2016).

The potential benefits of the model are uplifting. If only a few per cent reduction in mortality could be achieved, this would represent several tens of thousands of lives saved annually, worldwide (Fleischmann et al., 2015). In the last 10 to 15 years, attempts to develop novel treatments to reduce sepsis mortality have uniformly been unsuccessful (Angus et al., 2015b; Gordon et al., 2016; P. E. Marik, 2015; Ranieri et al., 2012). The use of computer decision support systems to better guide our current treatments and improve outcomes is therefore a much-needed approach. No such tool currently exists in research or clinical practice. The only related examples are represented by automatic algorithms that are advertised as being able to predict patient deterioration or the onset of sepsis (Hravnak et al., 2011; Nemati et al., 2018).

In the future, we hope to move this project towards the clinical environment for testing. We envision conducting real-time observational studies, comparing prospectively the output of our models to actual physicians but (initially) without altering clinical care. After collecting real patient and clinicians' data in a prospective fashion, we will replicate the evaluation methods described in this work: comparing the outcome of patients when the decisions of the AI and the clinician match or differ, estimating the WIS-based value of the clinicians' policy, etc. In a nutshell, early feasibility demonstration will require ethical clearance and the development of a secure prototype streaming patient data in real-time from an existing electronic health record software. We intend to conduct small-scale, pilot-like studies to gather user feedback and improve the design and integration of the tool into clinicians' workflow. "Closing the loop" (leaving the AI drive sepsis resuscitation autonomously) is not intended at this point, and several intermediate validation studies will be conducted.

A key challenge for the successful development of a clinical tool will be to maximise its acceptability by all stakeholders: patients, clinicians and regulators. Maximising the acceptability will require gaining the trust of these different groups, which will require demonstrating both the effectiveness and the safety of the technology.

Users embrace and adopt technologies such as navigation systems and home personal assistants because they deliver value. In healthcare, demonstrating value or effectiveness can take one of two forms: retrospective or prospective, the latter being considered much stronger in the hierarchy of evidence and is essential in the development of new drugs and devices (Frieden, 2017; Murad et al., 2016). To the best of our knowledge, no RCT has ever tested an RL-based decision support system in an acute clinical environment.

Demonstrating the safety could be achieved by following a few core principles. First of all, we'll ensure that data privacy will be preserved at all times, and that there is no risk of data privacy breach. The solution will comply with the General Data Protection Regulation (GDPR). Another important aspect of improving the safety and the accountability of our algorithm implies enhancing the transparency and the explainability of the AI decisions. The 'right to explanation' is a key part of achieving accountability and a significant aspect of the GDPR. This is a difficult task when using "big data" as the information is complex and so large that it is difficult to summarise. We have already made significant steps in this direction, as presented in the previous chapters (see Figure 61 for example). Finally, in our approach, we have put several safeguards in place, such as limiting the exploration capability of the AI agent to frequently observed actions taken by clinicians instead of allowing the agent to "free-roam". This seems a sensible step in early stage development to ensure safety. As the tool develops it may be possible to carefully expand the treatment strategies that the AI agent is able to suggest. Another safeguard is to leave the physician in charge and responsible for patient care. We do not envision allowing the AI solution to

directly affect the prescription of medications. Instead, the dose range suggested by the AI will be presented to the clinician in the form of a decision support system, which he will be free to follow or ignore. However, with more prospective testing and validation it may become feasible to allow the AI agent to automatically adjust infusion rates, with only supervision from clinicians.

Conclusion

It has been proposed that combining “machine learning software with the best human clinician “hardware” will permit delivery of care that outperforms what either can do alone” (Chen & Asch, 2017). In this work, we showed promising results about how machine learning, and in particular reinforcement learning, could be applied to solve a complex medical problem such as sepsis, and suggest individualised and clinically interpretable resuscitation strategies for sepsis. We hope that the work pioneered here will be reproduced by others and that we can continue to improve it, for the ultimate benefit of patients.

Acknowledgements

This project was funded by the Engineering and Physical Sciences Research Council and an Imperial College President's PhD Scholarship.

I am grateful to:

My family;

Sarah;

Anthony Gordon for his wholehearted support and unconditional tolerance to my hectic working style;

Aldo Faisal for enabling me to set a foot in the door of academia back in 2014;

Leo Celi for his great mentorship and for introducing me to the delights of J.P. Licks, kickboxing and the 5+5 challenge;

Omar Badawi and Philips Healthcare for providing access to the eRI data;

Finale Doshi-Velez and Omer Gottesman for their assistance with the methodology;

Farah Al-Beidh for the baked brownies;

Ed Palmer for his help with wrangling CCHIC;

Roger Mark and the Laboratory of Computational Physiology at MIT for their hospitality.

References

- Acheampong, A., & Vincent, J.-L. (2015). A positive fluid balance is an independent prognostic factor in patients with sepsis. *Critical Care*, *19*(1), 251. <https://doi.org/10.1186/s13054-015-0970-1>
- Aghabozorgi, S., Seyed Shirshorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. <https://doi.org/10.1016/j.is.2015.04.007>
- Akogul, S., Erisoglu, M., Akogul, S., & Erisoglu, M. (2017). An Approach for Determining the Number of Clusters in a Model-Based Cluster Analysis. *Entropy*, *19*(9), 452. <https://doi.org/10.3390/e19090452>
- Alagoz, O., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2004). The Optimal Timing of Living-Donor Liver Transplantation. *Management Science*, *50*(10), 1420–1430. <https://doi.org/10.1287/mnsc.1040.0287>
- Allen-Dicker, J. (2015, March 24). The Final Nail in Early Goal Directed Therapy’s Coffin? Retrieved March 4, 2016, from <http://blogs.nejm.org/now/index.php/the-final-nail-in-early-goal-directed-therapys-coffin/2015/03/24/>
- Alsous, F., Khamiees, M., DeGirolamo, A., Amoateng-Adjepong, Y., & Manthous, C. A. (2000). Negative fluid balance predicts survival in patients with septic shock: a retrospective pilot study. *Chest*, *117*(6), 1749–1754.
- Angus, D. C., Barnato, A. E., Bell, D., Bellomo, R., Chong, C.-R., Coats, T. J., ... Young, J. D. (2015a). A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. *Intensive Care Medicine*, *41*(9), 1549–1560. <https://doi.org/10.1007/s00134-015-3822-1>
- Angus, D. C., Barnato, A. E., Bell, D., Bellomo, R., Chong, C.-R., Coats, T. J., ... Young, J. D. (2015b). A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. *Intensive Care Medicine*, *41*(9), 1549–1560. <https://doi.org/10.1007/s00134-015-3822-1>
- Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., & Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine*, *29*(7), 1303–1310.
- ARDSNet. (2000). Ventilation with Lower Tidal Volumes as Compared with Traditional Tidal Volumes for Acute Lung Injury and the Acute Respiratory Distress Syndrome. *New England Journal of Medicine*, *342*(18), 1301–1308. <https://doi.org/10.1056/NEJM200005043421801>
- Artero, A., Zaragoza, R., & Miguel, J. (2012). Epidemiology of Severe Sepsis and Septic Shock. In R. Fernandez (Ed.), *Severe Sepsis and Septic Shock - Understanding a Serious Killer*. InTech.
- Arthur, D., & Vassilvitskii, S. (2007). K-means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. Retrieved from <http://dl.acm.org/citation.cfm?id=1283383.1283494>
- Asfar, P., Meziani, F., Hamel, J.-F., Grelon, F., Megarbane, B., Anguel, N., ... SEPSISPAM Investigators. (2014). High versus low blood-pressure target in patients with septic shock. *The New England Journal of Medicine*, *370*(17), 1583–1593. <https://doi.org/10.1056/NEJMoa1312173>

- Avni, T., Lador, A., Lev, S., Leibovici, L., Paul, M., & Grossman, A. (2015). Vasopressors for the Treatment of Septic Shock: Systematic Review and Meta-Analysis. *PLOS ONE*, *10*(8), e0129305. <https://doi.org/10.1371/journal.pone.0129305>
- Aytar, Y., Pfaff, T., Budden, D., Paine, T. L., Wang, Z., & de Freitas, N. (2018). Playing hard exploration games by watching YouTube. *ArXiv:1805.11592 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1805.11592>
- Bai, X., Yu, W., Ji, W., Lin, Z., Tan, S., Duan, K., ... Li, N. (2014). Early versus delayed administration of norepinephrine in patients with septic shock. *Critical Care (London, England)*, *18*(5), 532. <https://doi.org/10.1186/s13054-014-0532-y>
- Beam, A. L., & Kohane, I. S. (2016). Translating Artificial Intelligence Into Clinical Care. *JAMA*, *316*(22), 2368–2369. <https://doi.org/10.1001/jama.2016.17217>
- Beck, V., Chateau, D., Bryson, G. L., Pisipati, A., Zanotti, S., Parrillo, J. E., & Kumar, A. (2014). Timing of vasopressor initiation and mortality in septic shock: a cohort study. *Critical Care*, *18*(3), R97. <https://doi.org/10.1186/cc13868>
- Bellamy, M. C. (2006). Wet, dry or something else? *British Journal of Anaesthesia*, *97*(6), 755–757. <https://doi.org/10.1093/bja/ael290>
- Beloncle, F., Lerolle, N., Radermacher, P., & Asfar, P. (2013). Target blood pressure in sepsis: between a rock and a hard place. *Critical Care*, *17*(2), 126. <https://doi.org/10.1186/cc12543>
- Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, *57*(1), 9–19. <https://doi.org/10.1016/j.artmed.2012.12.003>
- Bishop, C. (2007). *Pattern Recognition and Machine Learning*. New York: Springer.
- Bock, A. S., & Fine, I. (2014). Anatomical and functional plasticity in early blind individuals and the mixture of experts architecture. *Frontiers in Human Neuroscience*, *8*, 971. <https://doi.org/10.3389/fnhum.2014.00971>
- Bonate, P. L. (2001). A brief introduction to Monte Carlo simulation. *Clinical Pharmacokinetics*, *40*(1), 15–22. <https://doi.org/10.2165/00003088-200140010-00002>
- Bosurgi, R. (2015). Sepsis: a need for new solutions. *The Lancet Infectious Diseases*, *15*(5), 498–499. [https://doi.org/10.1016/S1473-3099\(15\)70030-7](https://doi.org/10.1016/S1473-3099(15)70030-7)
- Bothe, M. K., Dickens, L., Reichel, K., Tellmann, A., Ellger, B., Westphal, M., & Faisal, A. A. (2013). The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Review of Medical Devices*, *10*(5), 661–673. <https://doi.org/10.1586/17434440.2013.827515>
- Bouza, C., López-Cuadrado, T., Saz-Parkinson, Z., & Amate-Blanco, J. M. (2014). Epidemiology and recent trends of severe sepsis in Spain: a nationwide population-based analysis (2006–2011). *BMC Infectious Diseases*, *14*, 3863. <https://doi.org/10.1186/s12879-014-0717-7>
- Boyd, J. H., Forbes, J., Nakada, T., Walley, K. R., & Russell, J. A. (2011). Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. *Critical Care Medicine*, *39*(2), 259–265. <https://doi.org/10.1097/CCM.0b013e3181feeb15>

- Brown, S. M., Lanspa, M. J., Jones, J. P., Kuttler, K. G., Li, Y., Carlson, R., ... Morris, A. H. (2013). Survival After Shock Requiring High-Dose Vasopressor Therapy. *Chest*, *143*(3), 664–671. <https://doi.org/10.1378/chest.12-1106>
- Byrne, L., & Haren, F. (2017). Fluid resuscitation in human sepsis: Time to rewrite history? *Annals of Intensive Care*, *7*(1), 4. <https://doi.org/10.1186/s13613-016-0231-8>
- Celi, Leo A., Csete, M., & Stone, D. (2014). Optimal data systems: the future of clinical predictions and decision support. *Current Opinion in Critical Care*, *20*(5), 573–580. <https://doi.org/10.1097/MCC.0000000000000137>
- Celi, Leo A., Mark, R. G., Stone, D. J., & Montgomery, R. A. (2013). “Big Data” in the Intensive Care Unit. Closing the Data Loop. *American Journal of Respiratory and Critical Care Medicine*, *187*(11), 1157–1160. <https://doi.org/10.1164/rccm.201212-2311ED>
- Celi, Leo Anthony, Hinske, L. C., Alterovitz, G., & Szolovits, P. (2008). An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study. *Critical Care*, *12*(6), R151. <https://doi.org/10.1186/cc7140>
- Celi, Leo Anthony, Zimolzak, A. J., & Stone, D. J. (2014). Dynamic Clinical Data Mining: Search Engine-Based Decision Support. *JMIR Medical Informatics*, *2*(1), e13. <https://doi.org/10.2196/medinform.3110>
- Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383.
- Chen, J. H., & Asch, S. M. (2017). Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. *New England Journal of Medicine*, *376*(26), 2507–2509. <https://doi.org/10.1056/NEJMp1702071>
- Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M. C., & Finkelstein, S. N. (2013). Missing data in medical databases: impute, delete or classify? *Artificial Intelligence in Medicine*, *58*(1), 63–72. <https://doi.org/10.1016/j.artmed.2013.01.003>
- Cohen, J., Vincent, J.-L., Adhikari, N. K. J., Machado, F. R., Angus, D. C., Calandra, T., ... Pelfrene, E. (2015). Sepsis: a roadmap for future research. *The Lancet Infectious Diseases*, *15*(5), 581–614. [https://doi.org/10.1016/S1473-3099\(15\)70112-X](https://doi.org/10.1016/S1473-3099(15)70112-X)
- Corrêa, T. D., Jakob, S. M., & Takala, J. (2015). Arterial blood pressure targets in septic shock: is it time to move to an individualized approach? *Critical Care*, *19*(1), 264. <https://doi.org/10.1186/s13054-015-0958-x>
- D’Agostino RB. (2007). Estimating treatment effects using observational data. *JAMA*, *297*(3), 314–316. <https://doi.org/10.1001/jama.297.3.314>
- Daniel M Faissol, P. M. G. (2007). Timing of Testing and Treatment of Hepatitis C and other Diseases. *Inform Journal on Computing - INFORMS*.
- Daskalaki, E., Diem, P., & Mougiakakou, S. G. (2016). Model-Free Machine Learning in Biomedicine: Feasibility Study in Type 1 Diabetes. *PLoS ONE*, *11*(7). <https://doi.org/10.1371/journal.pone.0158722>
- de Oliveira, F. S. V., Freitas, F. G. R., Ferreira, E. M., de Castro, I., Bafi, A. T., de Azevedo, L. C. P., & Machado, F. R. (2015). Positive fluid balance as a prognostic factor for mortality and acute kidney

injury in severe sepsis and septic shock. *Journal of Critical Care*, 30(1), 97–101. <https://doi.org/10.1016/j.jcrc.2014.09.002>

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., ... Surviving Sepsis Campaign Guidelines Committee including The Pediatric Subgroup. (2013). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Medicine*, 39(2), 165–228. <https://doi.org/10.1007/s00134-012-2769-8>

Denton, B. T., Kurt, M., Shah, N. D., Bryant, S. C., & Smith, S. A. (2009). Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 29(3), 351–367. <https://doi.org/10.1177/0272989X08329462>

Dombrovskiy, V. Y., Martin, A. A., Sunderram, J., & Paz, H. L. (2007). Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: a trend analysis from 1993 to 2003. *Critical Care Medicine*, 35(5), 1244–1250. <https://doi.org/10.1097/01.CCM.0000261890.41311.E9>

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. Retrieved from <https://arxiv.org/abs/1702.08608>

Dowling, H. F. (1972). Frustration and foundation. Management of pneumonia before antibiotics. *JAMA: The Journal of the American Medical Association*, 220(10), 1341–1345.

Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical Care*, 36(1), 8–27.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

Fleischmann, C., Scherag, A., Adhikari, N. K. J., Hartog, C. S., Tsaganos, T., Schlattmann, P., ... Reinhart, K. (2015). Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *American Journal of Respiratory and Critical Care Medicine*, 193(3), 259–272. <https://doi.org/10.1164/rccm.201504-0781OC>

Foss, A. H., & Markatou, M. (2018). kamila: Clustering Mixed-Type Data in R and Hadoop. *Journal of Statistical Software*, 83(1), 1–44. <https://doi.org/10.18637/jss.v083.i13>

Frieden, T. R. (2017). Evidence for Health Decision Making — Beyond Randomized, Controlled Trials. *New England Journal of Medicine*, 377(5), 465–475. <https://doi.org/10.1056/NEJMra1614394>

Gershman, S. J. (2017). Reinforcement Learning and Causal Models. In *The Oxford Handbook of Causal Reasoning* (p. 542). Oxford University Press.

Ghassemi, M., Celi, L. A., & Stone, D. J. (2015). State of the art review: the data revolution in critical care. *Critical Care*, 19(1), 118. <https://doi.org/10.1186/s13054-015-0801-4>

Ghassemi, M. M., Richter, S. E., Eche, I. M., Chen, T. W., Danziger, J., & Celi, L. A. (2014). A data-driven approach to optimized medication dosing: a focus on heparin. *Intensive Care Medicine*, 40(9), 1332–1339. <https://doi.org/10.1007/s00134-014-3406-5>

Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., & Szolovits, P. (2014). Unfolding Physiological State: Mortality Modelling in Intensive Care Units. In *Proceedings of*

the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 75–84). New York, NY, USA: ACM. <https://doi.org/10.1145/2623330.2623742>

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), E215-220.

Goodman, B., & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a “right to explanation.” *ArXiv:1606.08813 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1606.08813>

Gordon, A. C., Perkins, G. D., Singer, M., McAuley, D. F., Orme, R. M. L., Santhakumaran, S., ... Ashby, D. (2016). Levosimendan for the Prevention of Acute Organ Dysfunction in Sepsis. *New England Journal of Medicine*, *375*(17), 1638–1648. <https://doi.org/10.1056/NEJMoa1609409>

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine*, *25*(1), 16. <https://doi.org/10.1038/s41591-018-0310-5>

Gotts, J. E., & Matthay, M. A. (2016). Sepsis: pathophysiology and clinical management. *BMJ*, *353*, i1585. <https://doi.org/10.1136/bmj.i1585>

Group, T. D. R. (2017). Frequency of Evidence-Based Screening for Retinopathy in Type 1 Diabetes. *New England Journal of Medicine*, *376*(16), 1507–1516. <https://doi.org/10.1056/NEJMoa1612836>

Group, T. E. S. (2004). EPISEPSIS: a reappraisal of the epidemiology and outcome of severe sepsis in French intensive care units. *Intensive Care Medicine*, *30*(4), 580–588. <https://doi.org/10.1007/s00134-003-2121-4>

Gruber, S., & Laan, M. van der. (2009). Targeted Maximum Likelihood Estimation: A Gentle Introduction. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Retrieved from <http://biostats.bepress.com/ucbbiostat/paper252>

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, *316*(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>

Hanna, J. P., Stone, P., & Niekum, S. (2016). Bootstrapping with Models: Confidence Intervals for Off-Policy Evaluation. *ArXiv:1606.06126 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1606.06126>

Harris, S., Shi, S., Brealey, D., MacCallum, N. S., Denaxas, S., Perez-Suarez, D., ... Singer, M. (2018). Critical Care Health Informatics Collaborative (CCHIC): Data, tools and methods for reproducible research: A multi-centre UK intensive care database. *International Journal of Medical Informatics*, *112*, 82–89. <https://doi.org/10.1016/j.ijmedinf.2018.01.006>

Hex, N., Retzler, J., Bartlett, C., & Arber, M. (2017). *The Cost of Sepsis Care in the UK* (p. 46). York Health Economics Consortium.

Hjortrup, P. B., Haase, N., Bundgaard, H., Thomsen, S. L., Winding, R., Pettilä, V., ... Scandinavian Critical Care Trials Group. (2016). Restricting volumes of resuscitation fluid in adults with septic shock after initial management: the CLASSIC randomised, parallel-group, multicentre feasibility trial. *Intensive Care Medicine*, *42*(11), 1695–1705. <https://doi.org/10.1007/s00134-016-4500-7>

- Hravnak, M., DeVita, M. A., Clontz, A., Edwards, L., Valenta, C., & Pinsky, M. R. (2011). Cardiorespiratory instability before and after implementing an integrated monitoring system. *Critical Care Medicine*, 39(1), 65–72. <https://doi.org/10.1097/CCM.0b013e3181fb7b1c>
- Hu, X., & Xu, L. (2003). A Comparative Study of Several Cluster Number Selection Criteria. In J. Liu, Y. Cheung, & H. Yin (Eds.), *Intelligent Data Engineering and Automated Learning* (pp. 195–202). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-45080-1_27
- Hug, C. W. (2009). *Detecting hazardous intensive care patient episodes using real-time mortality models* (Thesis). Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/53290>
- Investigators, T. A., & Group, the A. C. T. (2014). Goal-Directed Resuscitation for Patients with Early Septic Shock. *New England Journal of Medicine*, 371(16), 1496–1506. <https://doi.org/10.1056/NEJMoa1404380>
- Investigators, T. P. (2014). A Randomized Trial of Protocol-Based Care for Early Septic Shock. *New England Journal of Medicine*, 370(18), 1683–1693. <https://doi.org/10.1056/NEJMoa1401602>
- Jagannatha, A., Thomas, P., & Yu, H. (2018). Towards High Confidence Off-Policy Reinforcement Learning for Clinical Applications. In *CausalML workshop* (p. 9).
- Jiang, N., & Li, L. (2015). Doubly Robust Off-policy Value Evaluation for Reinforcement Learning. *ArXiv:1511.03722 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1511.03722>
- Johnson, A. E. W., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D. A., & Clifford, G. D. (2016). Machine Learning and Decision Support in Critical Care. *Proceedings of the IEEE*, 104(2), 444–466. <https://doi.org/10.1109/JPROC.2015.2501978>
- Johnson, Alistair E. W., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, 160035. <https://doi.org/10.1038/sdata.2016.35>
- Jones, R. H. (2011). Bayesian information criterion for longitudinal and clustered data. *Statistics in Medicine*, 30(25), 3050–3056. <https://doi.org/10.1002/sim.4323>
- Kaukonen K, Bailey M, Suzuki S, Pilcher D, & Bellomo R. (2014). MOortality related to severe sepsis and septic shock among critically ill patients in australia and new zealand, 2000-2012. *JAMA*, 311(13), 1308–1316. <https://doi.org/10.1001/jama.2014.2637>
- Kelm, D. J., Perrin, J. T., Cartin-Ceba, R., Gajic, O., Schenck, L., & Kennedy, C. C. (2015). Fluid Overload in Patients with Severe Sepsis and Septic Shock Treated with Early-Goal Directed Therapy is Associated with Increased Acute Need for Fluid-Related Medical Interventions and Hospital Death. *Shock (Augusta, Ga.)*, 43(1), 68–73. <https://doi.org/10.1097/SHK.0000000000000268>
- Kipnis, E., & Vallet, B. (2010). Early norepinephrine resuscitation of life-threatening hypotensive septic shock: it can do the job, but at what cost? *Critical Care*, 14(6), 450. <https://doi.org/10.1186/cc9299>
- Kormushev, P., Calinon, S., & Caldwell, D. G. (2013). Reinforcement Learning in Robotics: Applications and Real-World Challenges. *Robotics*, 2(3), 122–148. <https://doi.org/10.3390/robotics2030122>

- Kreke, J. E. (2007). Modeling Disease Management Decisions for Patients with Pneumonia-related Sepsis [University of Pittsburgh ETD]. Retrieved from <http://d-scholarship.pitt.edu/8143/>
- Kshetri, K. B. (2013). Modelling Patient States in Intensive Care Patients. *MIT MEng Thesis*. Retrieved from http://groups.csail.mit.edu/medg/ftp/kshetri/Kshetri_MEng.pdf
- Lamontagne, F., Meade, M. O., Hébert, P. C., Asfar, P., Lauzier, F., Seely, A. J. E., ... Canadian Critical Care Trials Group. (2016). Higher versus lower blood pressure targets for vasopressor therapy in shock: a multicentre pilot randomized controlled trial. *Intensive Care Medicine*, *42*(4), 542–550. <https://doi.org/10.1007/s00134-016-4237-3>
- Lawrence, N. D. (2017). Data Readiness Levels. *ArXiv:1705.02245 [Cs]*. Retrieved from <http://arxiv.org/abs/1705.02245>
- Levy, M. M., Evans, L. E., & Rhodes, A. (2018). The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Medicine*, *44*(6), 925–928. <https://doi.org/10.1007/s00134-018-5085-0>
- Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., ... SCCM/ESICM/ACCP/ATS/SIS. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine*, *31*(4), 1250–1256. <https://doi.org/10.1097/01.CCM.0000050454.01978.3B>
- Lewis, S. R., Pritchard, M. W., Evans, D. J., Butler, A. R., Alderson, P., Smith, A. F., & Roberts, I. (2018). Colloids versus crystalloids for fluid resuscitation in critically ill people. *The Cochrane Database of Systematic Reviews*, *8*, CD000567. <https://doi.org/10.1002/14651858.CD000567.pub7>
- Li, L., Komorowski, M., & Faisal, A. A. (2018). The Actor Search Tree Critic (ASTC) for Off-Policy POMDP Learning in Medical Decision Making. *ArXiv:1805.11548 [Cs]*. Retrieved from <http://arxiv.org/abs/1805.11548>
- Littman, M. L. (2015). Reinforcement learning improves behaviour from evaluative feedback. *Nature*, *521*(7553), 445–451. <https://doi.org/10.1038/nature14540>
- Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.
- Mackenzie, D. C., & Noble, V. E. (2014). Assessing volume status and fluid responsiveness in the emergency department. *Clinical and Experimental Emergency Medicine*, *1*(2), 67–77. <https://doi.org/10.15441/ceem.14.040>
- Maillart, L. M., Ivy, J. S., Ransom, S., & Diehl, K. (2008). Assessing Dynamic Breast Cancer Screening Policies. *Operations Research*, *56*(6), 1411–1427. <https://doi.org/10.1287/opre.1080.0614>
- Malbrain, M. L. N. G., Marik, P. E., Witters, I., Cordemans, C., Kirkpatrick, A. W., Roberts, D. J., & Van Regenmortel, N. (2014). Fluid overload, de-resuscitation, and outcomes in critically ill or injured patients: a systematic review with suggestions for clinical practice. *Anaesthesiology Intensive Therapy*, *46*(5), 361–380. <https://doi.org/10.5603/AIT.2014.0060>
- Maldonado, G., & Greenland, S. (2002). Estimating causal effects. *International Journal of Epidemiology*, *31*(2), 422–429. <https://doi.org/10.1093/ije/31.2.422>
- Marik, P., & Bellomo, R. (2016). A rational approach to fluid therapy in sepsis. *British Journal of Anaesthesia*, *116*(3), 339–349. <https://doi.org/10.1093/bja/aev349>

- Marik, P. E. (2015). The demise of early goal-directed therapy for severe sepsis and septic shock. *Acta Anaesthesiologica Scandinavica*, 59(5), 561–567. <https://doi.org/10.1111/aas.12479>
- Marik, Paul E. (2014). Early management of severe sepsis: concepts and controversies. *Chest*, 145(6), 1407–1418. <https://doi.org/10.1378/chest.13-2104>
- Martin, G. S., Mannino, D. M., Eaton, S., & Moss, M. (2003). The Epidemiology of Sepsis in the United States from 1979 through 2000. *New England Journal of Medicine*, 348(16), 1546–1554. <https://doi.org/10.1056/NEJMoa022139>
- Mayr, F. B., Yende, S., & Angus, D. C. (2014). Epidemiology of severe sepsis. *Virulence*, 5(1), 4–11. <https://doi.org/10.4161/viru.27372>
- Micek, S. T., McEvoy, C., McKenzie, M., Hampton, N., Doherty, J. A., & Kollef, M. H. (2013). Fluid balance and cardiac function in septic shock as predictors of hospital mortality. *Critical Care*, 17(5), R246. <https://doi.org/10.1186/cc13072>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. Retrieved from <http://arxiv.org/abs/1301.3781>
- Miller, T. (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. *ArXiv:1706.07269 [Cs]*. Retrieved from <http://arxiv.org/abs/1706.07269>
- Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports*, 6, 26094. <https://doi.org/10.1038/srep26094>
- Misset, B., Nakache, D., Vesin, A., Darmon, M., Garrouste-Orgeas, M., Mourvillier, B., ... \$author.lastName, \$author firstName. (2008). Reliability of diagnostic coding in intensive care patients. *Critical Care*, 12(4), R95. <https://doi.org/10.1186/cc6969>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Moore, B. L., Pyeatt, L. D., Kulkarni, V., Panousis, P., Padrez, K., & Doufas, A. G. (2014). Reinforcement Learning for Closed-Loop Propofol Anesthesia: A Study in Human Volunteers. *Journal of Machine Learning Research*, 15, 655–696.
- Mouncey, P. R., Osborn, T. M., Power, G. S., Harrison, D. A., Sadique, M. Z., Grieve, R. D., ... Rowan, K. M. (2015). Trial of Early, Goal-Directed Resuscitation for Septic Shock. *New England Journal of Medicine*, 372(14), 1301–1311. <https://doi.org/10.1056/NEJMoa1500896>
- Munos, R., Stepleton, T., Harutyunyan, A., & Bellemare, M. G. (2016). Safe and Efficient Off-Policy Reinforcement Learning. *ArXiv:1606.02647 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1606.02647>
- Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine*, ebmmed-2016-110401. <https://doi.org/10.1136/ebmed-2016-110401>
- Murdoch TB, & Detsky AS. (2013). THE inevitable application of big data to health care. *JAMA*, 309(13), 1351–1352. <https://doi.org/10.1001/jama.2013.393>
- Myburgh, J. A., Finfer, S., Bellomo, R., Billot, L., Cass, A., Gattas, D., ... Australian and New Zealand Intensive Care Society Clinical Trials Group. (2012). Hydroxyethyl starch or saline for fluid

resuscitation in intensive care. *The New England Journal of Medicine*, 367(20), 1901–1911. <https://doi.org/10.1056/NEJMoa1209759>

Narayanan, M., Chen, E., He, J., Kim, B., Gershman, S., & Doshi-Velez, F. (2018). How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation. *ArXiv:1802.00682 [Cs]*. Retrieved from <http://arxiv.org/abs/1802.00682>

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018, April). An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU [Text]. <https://doi.org/info:doi/10.1097/CCM.0000000000002936>

Norris, J. R. (1997). Discrete-time Markov chains. In *Markov Chains*. Cambridge University Press. Retrieved from <http://dx.doi.org/10.1017/CBO9780511810633.003>

Office of the National Coordinator for Health Information Technology. (2016, May). Non-federal Acute Care Hospital Electronic Health Record Adoption. Retrieved August 27, 2018, from </quickstats/pages/FIG-Hospital-EHR-Adoption.php>

Parbhoo, S., Bogojeska, J., Zazzi, M., Roth, V., & Doshi-Velez, F. (2017). Combining Kernel and Model Based Learning for HIV Therapy Selection. *AMIA Summits on Translational Science Proceedings, 2017*, 239–248.

Prasad, N., Cheng, L.-F., Chivers, C., Draugelis, M., & Engelhardt, B. E. (2017). A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units. *ArXiv:1704.06300 [Cs]*. Retrieved from <http://arxiv.org/abs/1704.06300>

Precup, D., Sutton, R. S., & Singh, S. P. (2000). Eligibility Traces for Off-Policy Policy Evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 759–766). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=645529.658134>

Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.

Raith, E. P., Udy, A. A., Bailey, M., McGloughlin, S., MacIsaac, C., Bellomo, R., ... Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE). (2017). Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. *JAMA*, 317(3), 290–300. <https://doi.org/10.1001/jama.2016.20328>

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *Npj Digital Medicine*, 1(1), 18. <https://doi.org/10.1038/s41746-018-0029-1>

Ranieri, V. M., Thompson, B. T., Barie, P. S., Dhainaut, J.-F., Douglas, I. S., Finfer, S., ... Williams, M. D. (2012). Drotrecogin Alfa (Activated) in Adults with Septic Shock. *New England Journal of Medicine*, 366(22), 2055–2064. <https://doi.org/10.1056/NEJMoa1202290>

Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., ... Klompas, M. (2017). Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA*, 318(13), 1241–1249. <https://doi.org/10.1001/jama.2017.13836>

- Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., ... Dellinger, R. P. (2017). Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Medicine*, 1–74. <https://doi.org/10.1007/s00134-017-4683-6>
- Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., ... Tomlanovich, M. (2001). Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock. *New England Journal of Medicine*, 345(19), 1368–1377. <https://doi.org/10.1056/NEJMoa010307>
- Rosner, M. H., Ostermann, M., Murugan, R., Prowle, J. R., Ronco, C., Kellum, J. A., ... Shaw, A. D. (2014). Indications and management of mechanical fluid removal in critical illness. *British Journal of Anaesthesia*, aeu297. <https://doi.org/10.1093/bja/aeu297>
- Russell, J. A., Walley, K. R., Singer, J., Gordon, A. C., Hébert, P. C., Cooper, D. J., ... Ayers, D. (2008). Vasopressin versus Norepinephrine Infusion in Patients with Septic Shock. *New England Journal of Medicine*, 358(9), 877–887. <https://doi.org/10.1056/NEJMoa067373>
- Sakr, S., Elshawi, R., Ahmed, A. M., Qureshi, W. T., Brawner, C. A., Keteyian, S. J., ... Al-Mallah, M. H. (2017). Comparison of machine learning techniques to predict all-cause mortality using fitness data: the Henry ford exercise testing (FIT) project. *BMC Medical Informatics and Decision Making*, 17(1), 174. <https://doi.org/10.1186/s12911-017-0566-6>
- Salgado, C. M., Azevedo, C., Proença, H., & Vieira, S. M. (2016). Missing Data. In *Secondary Analysis of Electronic Health Records* (pp. 143–162). Springer, Cham. https://doi.org/10.1007/978-3-319-43742-2_13
- Sato, R., & Nasu, M. (2015). A review of sepsis-induced cardiomyopathy. *Journal of Intensive Care*, 3(1), 48. <https://doi.org/10.1186/s40560-015-0112-5>
- Schaefer, A. J., Bailey, M. D., Shechter, S. M., & Roberts, M. S. (2005). Modeling Medical Treatment Using Markov Decision Processes. In M. L. Brandeau, F. Sainfort, & W. P. Pierskalla (Eds.), *Operations Research and Health Care* (pp. 593–612). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/1-4020-8066-2_23
- Semler, M. W., & Rice, T. W. (2016). Sepsis Resuscitation: Fluid Choice and Dose. *Clinics in Chest Medicine*, 37(2), 241–250. <https://doi.org/10.1016/j.ccm.2016.01.007>
- Severs, D., Hoorn, E. J., & Rookmaaker, M. B. (2015). A critical appraisal of intravenous fluids: from the physiological basis to clinical evidence. *Nephrology Dialysis Transplantation*, 30(2), 178–187. <https://doi.org/10.1093/ndt/gfu005>
- Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., ... Levy, M. M. (2017). Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine*, 376(23), 2235–2244. <https://doi.org/10.1056/NEJMoa1703058>
- Seymour CW, Liu VX, Iwashyna TJ, & et al. (2016). Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8), 762–774. <https://doi.org/10.1001/jama.2016.0288>
- Shankar-Hari, M., Harrison, D. A., Rubenfeld, G. D., Rowan, K., & Myles, P. (2017). Epidemiology of sepsis and septic shock in critical care units: comparison between sepsis-2 and sepsis-3 populations using a national critical care database. *BJA: British Journal of Anaesthesia*, 119(4), 626–636. <https://doi.org/10.1093/bja/aex234>

- Shechter, S. M., Bailey, M. D., Schaefer, A. J., & Roberts, M. S. (2008). The Optimal Time to Initiate HIV Therapy Under Ordered Health States. *Operations Research*, 56(1), 20–33. <https://doi.org/10.1287/opre.1070.0480>
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>
- Singer M, Deutschman CS, Seymour C, & et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>
- Sirvent, J.-M., Ferri, C., Baró, A., Murcia, C., & Lorenzo, C. (2015). Fluid balance in sepsis and septic shock as a determining factor of mortality. *The American Journal of Emergency Medicine*, 33(2), 186–189. <https://doi.org/10.1016/j.ajem.2014.11.016>
- Smith, Q. W., Street, R. L., Volk, R. J., & Fordis, M. (2013). Differing Levels of Clinical Evidence: Exploring Communication Challenges in Shared Decision Making. *Medical Care Research and Review*, 70(1_suppl), 3S-13S. <https://doi.org/10.1177/1077558712468491>
- Spaan, M. T. J. (2012). Partially Observable Markov Decision Processes. In M. Wiering & M. van Otterlo (Eds.), *Reinforcement Learning: State-of-the-Art* (pp. 387–414). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-27645-3_12
- Stevenson, E. K., Rubenstein, A. R., Radin, G. T., Wiener, R. S., & Walkey, A. J. (2014). Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis*. *Critical Care Medicine*, 42(3), 625–631. <https://doi.org/10.1097/CCM.0000000000000026>
- Subramanian, S., Yilmaz, M., Rehman, A., Hubmayr, R. D., Afessa, B., & Gajic, O. (2008). Liberal vs. conservative vasopressor use to maintain mean arterial blood pressure during resuscitation of septic shock: an observational study. *Intensive Care Medicine*, 34(1), 157–162. <https://doi.org/10.1007/s00134-007-0862-1>
- Suk, H.-I., Lee, S.-W., Shen, D., & The Alzheimer’s Disease Neuroimaging Initiative. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function*, 220(2), 841–859. <https://doi.org/10.1007/s00429-013-0687-3>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). Cambridge, Mass: MIT Press.
- The PRISM Investigators. (2017). Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. *New England Journal of Medicine*, 376(23), 2223–2234. <https://doi.org/10.1056/NEJMoal701380>
- Thomas, P. S., & Brunskill, E. (2016). Data-Efficient Off-Policy Policy Evaluation for Reinforcement Learning. *ArXiv:1604.00923 [Cs]*. Retrieved from <http://arxiv.org/abs/1604.00923>
- Thomas, P. S., Theodorou, G., & Ghavamzadeh, M. (2015). High-Confidence Off-Policy Evaluation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Thomas, P., Theodorou, G., & Ghavamzadeh, M. (2015). High Confidence Policy Improvement. In *PMLR* (pp. 2380–2388). Retrieved from <http://proceedings.mlr.press/v37/thomas15.html>
- Torio, C. M., & Andrews, R. M. (2013). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief #160. In *Healthcare Cost and Utilization Project (HCUP)*

Statistical Briefs. Rockville (MD): Agency for Health Care Policy and Research (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK169005/>

Tsoukalas, A., Albertson, T., & Tagkopoulos, I. (2015). From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. *JMIR Medical Informatics*, 3(1), e11. <https://doi.org/10.2196/medinform.3445>

Tutz, G., & Ramzan, S. (2015). Improved methods for the imputation of missing data by nearest neighbor methods. *Computational Statistics & Data Analysis*, 90, 84–99. <https://doi.org/10.1016/j.csda.2015.04.009>

Undurti, A., Geramifard, A., & How, J. P. (2011). Function Approximation for Continuous Constrained MDPs. Retrieved August 26, 2018, from /paper/Function-Approximation-for-Continuous-Constrained-Undurti-Geramifard/347f7d2764d665f4f645456b6ad7dbc8b73f9c68

Vincent, J. L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., ... Thijs, L. G. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Medicine*, 22(7), 707–710.

Vincent, J.-L. (2016). The Future of Critical Care Medicine: Integration and Personalization. *Critical Care Medicine*, 44(2), 386–389. <https://doi.org/10.1097/CCM.0000000000001530>

Vincent, J.-L., Sakr, Y., Sprung, C. L., Ranieri, V. M., Reinhart, K., Gerlach, H., ... Sepsis Occurrence in Acutely Ill Patients Investigators. (2006). Sepsis in European intensive care units: results of the SOAP study. *Critical Care Medicine*, 34(2), 344–353.

Waechter, J., Kumar, A., Lapinsky, S. E., Marshall, J., Dodek, P., Arabi, Y., ... Cooperative Antimicrobial Therapy of Septic Shock Database Research Group. (2014). Interaction between fluids and vasoactive agents on mortality in septic shock: a multicenter, observational study. *Critical Care Medicine*, 42(10), 2158–2168. <https://doi.org/10.1097/CCM.0000000000000520>

Welton, N. J., Sutton, A. J., Cooper, N., Abrams, K. R., & Ades, A. E. (2012). *Evidence Synthesis for Decision Making in Healthcare*. John Wiley & Sons.

Zhang, B., & An, B. (2018). Clustering time series based on dependence structure. *PLOS ONE*, 13(11), e0206753. <https://doi.org/10.1371/journal.pone.0206753>

Appendices

Code availability

The SQL and Matlab code to recreate the models and generate the figures have been deposited and are available at https://github.com/matthieukomorowski/AI_Clinician and <http://www.imperial.ac.uk/artificial-intelligence/research/healthcare/ai-clinician>

Research output as of February 2019

Peer-reviewed papers

1. O Gottesman, F Johansson, **M Komorowski**, A Faisal, D Sontag, F Doshi-Velez, LA Celi, “The promises and pitfalls of reinforcement learning in healthcare”, Nature Medicine, Jan 2019
2. S. Parbhoo, O Gottesman, AS Ross, **M Komorowski**, A Faisal, I Bon, V Roth, F Doshi-Velez. “Improving counterfactual reasoning with kernelised dynamic mixing models”, PLOS ONE, December 2018
3. **M Komorowski**, LA Celi, O. Badawi, AC Gordon, A Faisal, “The intensive care AI clinician learns optimal treatment strategies for sepsis.”, Nature Medicine, November 2018
4. R Deliberato, A Neto, **M Komorowski**, DJ Stone, S Ko, L Bulgarelli, C Rodrigues Ponzoni, R Carneiro de Freitas Chaves, LA Celi, AEW Johnson “An Evaluation of the Influence of Body Mass Index on Severity Scoring”, Crit Care Med, November 2018
5. DC Marshall; RJ Goodson; Y Xu; **M Komorowski**; M Maruthappu; J Shalhoub; JD Salciccioli, “Trends in Mortality from Pneumonia in the European Union: a Temporal Analysis of the European Detailed Mortality Database between 2001-2014” Respiratory Research, May 2018 19:81
6. R Deliberato, S Ko, **M Komorowski**, MA Armengol de la Hoz, M Frushicheva, J Raffa, AEW Johnson, L Celi, DS Stone, “Severity of illness scores may misclassify critically ill obese patients”, Crit Care Med, Nov 2017. DOI: 10.1097/CCM.0000000000002868
7. **M. Komorowski**, LA Celi “Will Artificial Intelligence Contribute to Overuse in Healthcare?” Crit Care Med, 45(5):912–913, May 2017. DOI: 10.1097/CCM.0000000000002351
8. La Celi, G Davidzon, Johnson AE, **Komorowski M**, Marshall DC, Nair SS, Phillips CT, Pollard TJ, Raffa JD, Salciccioli JD, Salgueiro FM, Stone DJ “Bridging the Health Data Divide” J Med Internet Res, Dec 2016;18(12):e325 DOI: 10.2196/jmir.6400

Book and book chapters

Co-editor of open access textbook: “Secondary Analysis of Electronic Health Records”, Springer International Publishing, Sept 2016. The book was downloaded 362,000 times as of 22/01/2019.

Co-author of chapters:

1. **M. Komorowski**, D. C. Marshall, J. D. Saliccioli, and Y. Crutain, “Exploratory Data Analysis” in Secondary Analysis of Electronic Health Records, Springer International Publishing, Sept 2016, pp. 185–203.
2. J. D. Saliccioli, Y. Crutain, **M. Komorowski**, and D. C. Marshall, “Sensitivity Analysis and Model Validation” in Secondary Analysis of Electronic Health Records, Springer International Publishing, Sept 2016, pp. 263–271.
3. **M. Komorowski** & J. Raffa, “Markov Models and Cost Effectiveness Analysis: Applications in Medical Research” in Secondary Analysis of Electronic Health Records, Springer International Publishing, Sept 2016, pp. 351–367.

Abstracts

1. A Raghu, **M Komorowski**, S Singh. “Model-Based Reinforcement Learning for Sepsis Treatment” NIPS ML4H Workshop, Montreal, Canada, Dec 2018
2. JD Saliccioli, **M Komorowski**, J Shalhoub, DC Marshall, P Clardy. “Global Trends in Sepsis Mortality: a temporal analysis of the WHO Mortality Database between 1985 and 2015”, ESICM, Paris, Oct 2018
3. A Raghu, O Gottesman, Y Liu, **M Komorowski**, A Faisal, F Doshi-Velez, E Brunskill. “Behaviour Policy Estimation in Off-Policy Policy Evaluation: Calibration Matters” FAIM Workshop on CausalML, Stockholm, July 2018
4. O Gottesman, Y Liu, A Raghu, **M Komorowski**, A Faisal, F Doshi-Velez, E Brunskill. “Representation Balancing MDPs for Off-Policy Policy Evaluation”, FAIM Workshop on CausalML, Stockholm, July 2018
5. O.Ren, A. Johnson, **M. Komorowski**, J. Aboab, Z. Shahn, D. Sow, R. Mark, L. Lehman. “Predicting and Understanding Unexpected Respiratory Decompensation in Critical Care”, the Sixth IEEE International Conference on Healthcare Informatics, New York, June 2018
6. A. Raghu, **M. Komorowski**, I. Ahmed, L. Celi, P. Szolovits, M. Ghassemi. “Deep Reinforcement Learning for Sepsis Treatment”, Transparent and Interpretable Machine Learning in Safety Critical Environments, Long Beach CA, Dec 2017

7. A. Raghu, M. Ghassemi, **M. Komorowski**, L. Celi, P. Szolovits. “Continuous State-Space Models for Optimal Sepsis Treatment - a Deep Reinforcement Learning Approach”, Machine Learning for Healthcare, Boston, April 2017
8. **M. Komorowski**, A. Gordon, A. Faisal, “Optimising the management of severe infections in intensive care with reinforcement learning”, AI in Medicine conference, Laguna Niguel, December 2016
9. **M. Komorowski**, “Clinician pitch: Data-driven techniques to optimise the management of septic shock”, NIPS ML4H workshop, Barcelona, December 2016
10. **M. Komorowski**, A. Gordon, A. Faisal, “A Markov Decision Process to suggest optimal treatment of severe infections in intensive care.”, NIPS ML4H workshop, Barcelona, December 2016
11. J. Saliccioli, P. Charlton, A. Hartley, **M. Komorowski**, D. Marshall, J. Shalhoub, M.C. Sykes, L.A. Celi “Lactate Rebound as an Independent Predictor of Mortality in Intensive Care”, ATS Conference, San Francisco, May 2016
12. **M. Komorowski** & A. Faisal “Towards Closed-Loop Mortality Prediction and Off-Policy Learning of Medical Decision Derived from Very Large Scale Intensive Care Unit Databases” [extended abstract], Multidisciplinary Conference on Reinforcement Learning and Decision Making, Edmonton, Alberta, Canada, June 7-10, 2015.
13. **M. Komorowski**, A. Faisal, “Visualising patients' dynamics in the ICU and predicting mortality in real-time using big data” [poster], International Symposium on Intensive Care and Emergency Medicine, Brussels, March 2015.

Reviewer activity

Ad-hoc Reviewer for the journals:

1. *Biotechnology Advances* (IF 9.85)
2. *Critical Care Medicine* (IF 7.44)
3. *British Journal of Anaesthesia* (IF 6.49)
4. *Critical Care* (IF 4.95)
5. *Value in Health* (IF 3.28)
6. *Journal of Critical Care* (IF 2.9)
7. *BMJ Open* (IF 2.37)
8. *Medicine* (IF 2.13)
9. *Aerospace Medicine and Human Performance* (IF 0.88)
10. *Anesthésie & Réanimation* (IF 0.84)

11. The publisher *Elsevier*.

Prizes and awards

1. Microsoft/Akshaya Patra/Asha Hackathon: 1st place for project “Using tech to reduce malnutrition in rural India” (Nov 2018)
2. Imperial College London - Student Academic Choice Award (March 2018)
3. DAT-ICU intensive care datathon, AP-HP & French Intensive Care Society, Paris (Jan 2018): 1st place with project “Improving sepsis treatment with patient similarity metrics”
4. 2nd Chinese PLA General Hospital – MIT datathon, Beijing, China (Nov 2017): 2nd place
5. Taipei Medical University-MIT hackathon, Taipei, Taiwan (Oct 2017): 1st place with project “Predicting fluid responsiveness with feed-forward deep neural networks”
6. Artificial Intelligence in Medicine Conference best abstract award in the category “Decision Support and Hospital Monitoring”, Laguna Niguel, CA (Dec 2016)
7. **Royal Society of Medicine**, Emergency Medicine Section, Research and Innovation first prize, London (July 2016). Value £500
8. Data Science London Open Healthcare Data Hack: 1st place in category machine learning, London (Dec 2014). Value £500

Invited talks

1. International Committee on Military Medicine Conference – Basel, Switzerland (May 2019)
2. International Symposium on Intensive Care and Emergency Medicine, Brussels (March 2019)
3. The Big Sick conference, Zermatt, Switzerland (Feb 2019)
4. ESICM Big Datatalk, Milan, Italy (Feb 2019)
5. Society for Technology in Anesthesia, Scottsdale, AZ (Jan 2019)
6. University of Pittsburgh Medical Center, Pittsburgh, MA (Jan 2019)
7. AIMed annual conference, Laguna Niguel, CA, USA (Dec 2018)
8. Intensive Care Society state of the art meeting, London (Dec 2018): “ The rise of the Machines” and Gold Medal competition “Optimizing sepsis treatment with machine learning”
9. WeARe conference, Paris (Dec 2018): “Improving sepsis resuscitation with machine learning”
10. Department of Intensive Care weekly Seminar, Argenteuil-Colombes hospital, Paris (Dec 2018): “Machine learning in sepsis research”
11. Holst symposium on Artificial Intelligence, Eindhoven University of Technology (Nov 2018)

12. European Society of Intensive Care Medicine, Paris (Oct 2018): “Improving sepsis treatment with machine learning”
13. AIMed Europe conference, London, UK (Sept 2018): “A comparative overview of the application & use of AI in medicine: Europe, Middle East and USA”
14. Department of Anaesthesia and Intensive Care weekly Seminar, Fort-de-France University Hospital (Aug 2018): “Machine learning in sepsis research”
15. AIMed breakfast briefing, The Wellcome Collection, London (July 2018): “How can artificial intelligence help the NHS?”
16. Taipei Medical University and AHMC Health Inc., Los Angeles, CA (March 2018): “Deep Intelligence to Advance Smart Care- Towards A Learning Health System”
17. The Big Sick conference, Zermatt, Switzerland (Feb 2018): “Healthcare applications of machine learning: the example of sepsis”
18. Royal Society of Medicine, London (Feb 2018): “Beyond the Surviving Sepsis Campaign: the Surviving Sepsis Computer.” – Winner of the Royal Society of Medicine first prize in category “Emergency Medicine, Section Research and Innovation”
19. AIMed annual conference, Laguna Niguel, CA, USA (Dec 2017): “Reinforcement Learning in Sepsis”
20. Chinese PLA General Hospital, Beijing, China (Nov 2017): “Reinforcement Learning in Intensive Care”
21. Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA (Sept 2017): “Healthcare applications of machine learning: the example of sepsis”
22. World Sepsis Day, Lens General Hospital, France (Sept 2017): “Artificial intelligence and the future of management of sepsis”
23. Intensive Care Department, Beth Israel Deaconess Medical Center, Boston, MA (Feb 2017): “Optimising medical decisions in critical care with reinforcement learning”

Oral Presentations for submitted abstracts

1. Imperial College London, Division of Surgery Research Afternoon (Dec 2017): “Optimising the management of sepsis in intensive care with reinforcement learning”
2. European Society of Intensive Care Medicine, Vienna (Sept 2017): “Severity of illness scores may misclassify critically ill obese patients”
3. Artificial Intelligence in Medicine conference, Laguna Niguel, USA (Dec 2016): “Optimising the management of severe infections in intensive care with reinforcement learning” – Winner of best abstract in the category “Decision Support and Hospital Monitoring”

4. NIPS Machine learning for healthcare workshop, Barcelona (Dec 2016): “A Markov decision process to suggest optimal treatment of severe infections in intensive care” and “Clinician pitch: the challenge of sepsis” – Winner of NIPS Machine Learning for Healthcare travel scholarship
5. Data Science London Open Healthcare Data Hack, London (Dec 2014) “Mortality prediction and medical decision support system using MIMIC-II data” – Winner of best project in the category “Machine Learning”

Leadership and teaching experience

- Sept 2018: Co-chair of AIMed Europe conference (400 attendees) and AIMed datathon (70 participants)
- Dec 2017 and 2018: Lecturing for Imperial College BSc Surgery and Anaesthetics
- Sept-Dec 2017: Co-instructor, Harvard course CS282 “Reinforcement Learning in Healthcare”
- Sept-Dec 2016 and 2017: Lecturer, MIT HST953 course “Secondary Analysis of Health Records”
- June 2016: Lecturing Imperial College MRes Biomedical Research (Anaesthetics, Pain Medicine and Intensive Care) on health data analytics, statistics and machine learning

Permissions to reuse

6/12/2018

RightsLink Printable License

BMJ PUBLISHING GROUP LTD. LICENSE TERMS AND CONDITIONS

Jun 12, 2018

This Agreement between Matthieu Komorowski ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

License Number	4366550185164
License date	Jun 12, 2018
Licensed Content Publisher	BMJ Publishing Group Ltd.
Licensed Content Publication	The BMJ
Licensed Content Title	Sepsis: pathophysiology and clinical management
Licensed Content Author	Jeffrey E Gotts, Michael A Matthay
Licensed Content Date	May 23, 2016
Licensed Content Volume	353
Type of Use	Dissertation/Thesis
Requestor type	Individual
Format	Electronic
Portion	Figure/table/extract
Number of figure/table/extracts	1
Description of figure/table/extracts	Figure 3 "Organ failure in a critically ill patient with septic shock from pneumococcal pneumonia. ARDS=acute respiratory distress syndrome"
Will you be translating?	No
Circulation/distribution	1000
Title of your thesis / dissertation	MACHINE LEARNING APPROACHES TO OPTIMISE THE MANAGEMENT OF PATIENTS WITH SEPSIS
Expected completion date	Aug 2018
Estimated size(pages)	100
Requestor Location	Matthieu Komorowski 285, rue de Lille Bailleul, Nord 59270 France Attn:
Publisher Tax ID	GB674738491
Billing Type	Invoice
Billing Address	Matthieu Komorowski 285, rue de Lille Bailleul, France 59270 Attn: Matthieu Komorowski
Total	0.00 GBP
Terms and Conditions	

BMJ Group Terms and Conditions for Permissions

**WOLTERS KLUWER HEALTH, INC. LICENSE
TERMS AND CONDITIONS**

Jun 12, 2018

This Agreement between Matthieu Komorowski ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	4366550397782
License date	Jun 12, 2018
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Critical Care Medicine
Licensed Content Title	Two Decades of Mortality Trends Among Patients With Severe Sepsis: A Comparative Meta-Analysis*.
Licensed Content Author	Stevenson, Elizabeth; MD, MS; Rubenstein, Amanda; Radin, Gregory; Wiener, Renda; Soylemez MD, MPH; Walkey, Allan; MD, MSc
Licensed Content Date	Mar 1, 2014
Licensed Content Volume	42
Licensed Content Issue	3
Type of Use	Dissertation/Thesis
Requestor type	Individual
STM publisher name	
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	Figure 2
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	MACHINE LEARNING APPROACHES TO OPTIMISE THE MANAGEMENT OF PATIENTS WITH SEPSIS
Expected completion date	Aug 2018
Estimated size(pages)	100
Requestor Location	Matthieu Komorowski 285, rue de Lille Bailleul, Nord 59270 France Attn:
Publisher Tax ID	EU826013006
Billing Type	Invoice
Billing Address	Matthieu Komorowski 285, rue de Lille Bailleul, France 59270 Attn: Matthieu Komorowski
Total	0.00 USD
Terms and Conditions	

**WOLTERS KLUWER HEALTH, INC. LICENSE
TERMS AND CONDITIONS**

Sep 22, 2018

This Agreement between Matthieu Komorowski ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	4367020320433
License date	Jun 13, 2018
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Critical Care Medicine
Licensed Content Title	Fluid resuscitation in septic shock: A positive fluid balance and elevated central venous pressure are associated with increased mortality*
Licensed Content Author	John H. Boyd, Jason Forbes, Taka-aki Nakada, et al
Licensed Content Date	Feb 1, 2011
Licensed Content Volume	39
Licensed Content Issue	2
Type of Use	Dissertation/Thesis
Requestor type	Individual
STM publisher name	
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	Figure 2
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	MACHINE LEARNING APPROACHES TO OPTIMISE THE MANAGEMENT OF PATIENTS WITH SEPSIS
Expected completion date	Aug 2018
Estimated size(pages)	100
Requestor Location	Matthieu Komorowski 285, rue de Lille Bailleul, Nord 59270 France Attn:
Publisher Tax ID	EU826013006
Billing Type	Invoice
Billing Address	Matthieu Komorowski 285, rue de Lille Bailleul, France 59270 Attn: Matthieu Komorowski
Total	0.00 USD
Terms and Conditions	

**SPRINGER NATURE LICENSE
TERMS AND CONDITIONS**

Sep 22, 2018

This Agreement between Matthieu Komorowski ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4367011317979
License date	Jun 13, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Intensive Care Medicine
Licensed Content Title	Fluid administration in severe sepsis and septic shock, patterns and outcomes: an analysis of a large national database
Licensed Content Author	Paul E. Marik, Walter T. Linde-Zwirble, Edward A. Bittner et al
Licensed Content Date	Jan 1, 2017
Licensed Content Volume	43
Licensed Content Issue	5
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	MACHINE LEARNING APPROACHES TO OPTIMISE THE MANAGEMENT OF PATIENTS WITH SEPSIS
Instructor name	n/a
Institution name	n/a
Expected presentation date	Aug 2018
Portions	Figure 3
Requestor Location	Matthieu Komorowski 285, rue de Lille Bailleul, Nord 59270 France Attn:
Billing Type	Invoice
Billing Address	Matthieu Komorowski 285, rue de Lille Bailleul, France 59270 Attn: Matthieu Komorowski
Total	0.00 USD
Terms and Conditions	