STOCHASTIC TREATMENT RECOMMENDATION WITH DEEP SURVIVAL DOSE RESPONSE FUNCTION (DEEPSDRF)

A PREPRINT

Jie Zhu*
Centre for Big Data Research in Health (CBDRH)
UNSW, Sydney
NSW, 2052, Australia
elliott.zhu@unsw.edu.au

Blanca Gallego
Centre for Big Data Research in Health (CBDRH)
UNSW, Sydney
NSW, 2052, Australia
b.gallego@unsw.edu.au

August 25, 2021

ABSTRACT

We propose a general formulation for stochastic treatment recommendation problems in settings with clinical survival data, which we call the Deep Survival Dose Response Function (DeepSDRF). That is, we consider the problem of learning the conditional average dose response (CADR) function solely from historical data in which unobserved factors (confounders) affect both observed treatment and time-to-event outcomes. The estimated treatment effect from DeepSDRF enables us to develop recommender algorithms with explanatory insights. We compared two recommender approaches based on random search and reinforcement learning and found similar performance in terms of patient outcome. We tested the DeepSDRF and the corresponding recommender on extensive simulation studies and two empirical databases: 1) the Clinical Practice Research Datalink (CPRD) and 2) the eICU Research Institute (eRI) database. To the best of our knowledge, this is the first time that confounders are taken into consideration for addressing the stochastic treatment effect with observational data in a medical context.

***K*eywords** Recommendation system · Causal Inference · Continuous Treatments · Survival Outcomes

1 Introduction

Continuous treatments or exposures (such as dose, duration, and frequency) arise very often in clinical studies. Importantly, such treatments lead to effects that are naturally described by curves (e.g., dose response curves) rather than scalars, as might be the case for binary treatments. Two major methodological challenges in continuous treatment settings are (1) to allow for a flexible estimation of the dose response curve (for example, to discover the underlying structure without imposing a priori shape restrictions), and (2) to properly adjust for high-dimensional confounders (i.e., pre-treatment covariates related to treatment assignment and outcome).

The application of data-adaptive models to healthcare has yielded great advancements in personalized medicine. The vast majority of these focus on diagnosing conditions or forecasting outcomes, yet few on treatment recommendations. Reinforcement learning has been proposed as a solution to figuring out the optimal policy in the context of clinical studies for treatment advice to individual patients, but there remain many challenges in learning and evaluating on primarily observational data. The most prominent is the problem of confounding bias, where unobserved factors (confounders) affect both observed treatment assignment (policy) and the patient outcome and result in biased estimation and recommendations.

In this study, our goal is to apply theories of statistical causal inference to provide debiased treatment recommendations to patients. Specifically, we focus on the observational setting, that is, the setting in which our recommender will propose treatment levels based on confounding-adjusted estimations of the treatment effect, and evaluate the patient's

*

outcome based on the historical data. This retrospective setting is common in healthcare applications, where it is impossible or unethical to experiment with alternative treatment strategies on patients.

The confounding bias is perilous when there is a significant lack of support in the examined cohorts, such as the lack of data on nontreated patients, where patients with critical conditions, despite receiving treatments, frequently experienced adverse outcomes. This leads to the earned policy suggesting no treatment or treatment that physicians would never do. And measures such as importance sampling and U-Curve share the same failure when evaluating the policy given there is no similar patient under different treatment conditions [1] (i.e., lack of overlapping). Our proposed algorithm incorporates the general propensity score approach (GPS) [2] to adjust the expected outcomes of patients in terms of survival curves for potential confounders. Despite the frequency of survival outcomes in biomedical research and the increased use of GPS, there are, to the best of our knowledge, no nonparametric studies applying GPS for estimating the effect of stochastic exposure on time-to-event outcomes.

Secondly, we propose a recommender based on the potential difference in survival curves given the alternative and the original treatment levels. Compare to binary mortality outcomes, the survival curves possess the merit of balancing the short-term improvement with long-term success, which is discussed in our previous work on estimating the survival treatment effect. [3, 4] For example, when monitoring sepsis patients, popular outcome choices would be biomarkers such as the periodic reduction in white blood cells or the terminal outcome such as death. The survival curves, on the other hand, provide the survival probability as a unified feedback to clinicians which incorporate the effect of short-term events on the long-term probability of survival. Two recommender algorithms based on random search and reinforcement learning were compared and found to have similar performance.

As most major breakthroughs in deep learning have been trained on years worth of simulated data. Clearly, it is infeasible to obtain a large amount of data for a specific treatment in an observational context. We validated our model, which we called the Deep Survival Dose Response Function (DeepSDRF), mainly through simulators. The performance of the corresponding recommenders was in addition tested in a case study based on pseudo treatments created from the Clinical Practice Research Datalink (CPRD). [5] We demonstrate the application of DeepSDRF in two datasets generated using the eICU Research Institute (eRI) database [6]. One for the continuous vasopressor dosage assignment problem of sepsis treatment in intensive care units (ICUs). Another is the health economic problem of the optimal timing to send patients from the ICUs to Step Down Units (SDUs). Across simulated and real datasets, we show that the proposed treatment effect estimator DeepSDRF, and its associated recommendations are robust to confounders and improve the patient outcome.

Our proposed work solves three major methodological challenges in continuous treatment settings: (1) to allow for a flexible estimation of the dose response curve (i.e., to discover the underlying structure without imposing a priori parametric restrictions), (2) to properly adjust for high-dimensional confounders (i.e., pre-treatment covariates related to treatment assignment and outcome), (3) to enable the estimation of the effect of stochastic treatments on survival outcomes. In Section 2, we discuss the methodology and the study design. We present the results in Section 3 and conclude with a discussion.

2 Method

2.1 Causal inference for stochastic interventions

Our major goal is to measure the effect of stochastic intervention on the time-to-event outcomes. This causal parameter will be used for providing causality based treatment recommendations in a dynamic clinical context to individual patients. Current approaches for estimating the effect of stochastic intervention are based on regression models (i.e., the dose response function) of continuous outcomes on covariates and treatments.[2, 7] However, this approach relies entirely on the correct specification of the outcome model and is sensitive to the curse of dimensionality by inheriting the rate of convergence of the outcome regression estimator. The alternative generalised propensity score model approaches [2, 8, 9] manage to correct the selection bias of the treatment assignment, but they still rely on the accurate choice of the treatment propensity model.

In contrast, doubly robust estimators [10] are based on modeling both the treatment and outcome processes and give consistent treatment effect estimations as long as one of these two data generating processes is accurately modeled. The doubly robust methods converge faster than their nuisance estimators (i.e., outcome and treatment models) when both models are consistently estimated; this makes them less sensitive to the curse of dimensionality and can allow for inference using flexible machine learning-based adjustment. However, standard semi-parametric doubly robust methods for estimating dose response functions rely on the parametric model of the effect curves, either by explicitly assuming a parametric dose response function [11, 12], or by projecting the true response function onto a parametric working

model.[13] Unfortunately, the first approach can lead to substantial bias under model misspecification, and the second can be of limited practical use if the working model is far away from the truth.

Recent work has extended semi-parametric doubly robust methods to more complicated nonparametric and high-dimensional settings. The work of Super-Learner[12] proposed an empirical risk minimization framework for estimator selection in causal inference problems, and particularly in the average treatment effect estimation on survival outcomes.[14] We applied this framework to estimate the conditional average treatment effect of binary interventions on survival outcomes with static covariates [3] and extended the model to time-varying covariates using recurrent neural subnetworks in the working paper of Causal Dynamic Survival (CDS) model. [4] In this work, we further relax the assumptions on binary treatment to estimate the effect of time-varying stochastic treatments on time-to-event outcomes.

We present a new approach based on the doubly robust dose response function but without parametric assumptions. Our method has a simple two-stage implementation that is fast and easy to use with standard software: in the first stage, a general propensity score (GPS) model is constructed based on the mapping between the treatment and covariates; and in the second stage, we regress the estimated propensity score and the treatment on the survival outcomes designed similar to our previous studies.[3, 4] Both regression can be conducted via off-the-shelf nonparametric machine learning tools. We provide the asymptotic results of our approach with extensive simulations. We also discuss a simple method for general propensity score estimation when the probability density function of the treatment assignment process is unknown. The method is validated via simulations and an empirical database and illustrated in two case studies which will be discussed later about the sepsis treatment in intensive care units (ICUs) and the decision of early discharge from ICUs to step down units (SDUs).

2.2 Conditional average dose response function for survival outcomes

With time-to-event outcomes, we suggest that the concept of the dose–response function requires modification. If one were to use the definition of the dose–response function used for continuous or binary outcomes, then the value of the dose–response function for a given value of exposure would denote the expected survival time under that value of the stochastic exposure. There are two limitations to this approach. The first limitation is that it can be difficult to estimate the mean survival time in the presence of a moderate to high degree of censoring. The second limitation is that differences in survival time are not quantified by a standardised measure of treatment effect. Instead, differences in survival are often quantified using differences in survival curves. For these two reasons, we propose that the dose–response function be modified so that the dose–response function for a given value of the stochastic exposure denotes the survival function if all subjects in the sample were to receive the given value of the exposure. Formally, suppose we observe a sample $\mathcal{O}$ of n independent observations generated from an unknown distribution $\mathcal{P}_0$:

$$\mathcal{O} := \big(X_i(t), Y_i(t), A_i(t), t_i = \min(t_{s,i}, t_{c,i})\big), i \ = 1, 2, \ldots, n$$

where $X_i(t) = (X_{i,1}(t), X_{i,2}(t), \ldots, X_{i,d}(t)), d = 1, 2, \ldots, D$ are baseline covariates at time t, $t = 1, 2, \ldots, \Theta$, with Θ being the maximum follow-up time of the study; $A_i(t)$ is the treatment condition at time t; $Y_i(t)$ denotes the outcome at time t, $Y_i = 1$ if i experienced an event and $Y_i = 0$ otherwise; t_i is determined by the event or censor time, $t_{s,i}$ or $t_{c,i}$, whichever happened first. For simplicity, we drop individual indicator i in the sequel and assume A are defined on a common probability space, that A is continuously distributed with respect to Lebesgue measure on $\mathcal{A}$ and that Y is a well-defined random variable (this requires that the random function $Y(\cdot)$ be suitably measurable). One could speak of a dose–response surface measured in terms of the hazard rate at time t over a history of treatment levels $a \in \mathcal{A}$ as:

$$h(t, a) \ := \ Pr(Y(t) = 1 \mid \overline{A}(t, u) = a, \overline{X}(t, u)),$$

which is the probability of experiencing an event in the interval $(t-1, t]$ for individual i given the history of treatments and covariates from $t-u$ to $t-1$ with u being the length of the observation history. Note in the observation space $\mathcal{O}$, for each patient the trajectory of $A_i(t)$ is stochastic, and we denote the level at each time point as $A(t) = a_t$. While to calculate the dose response surface, we assume the potential treatment level would be fixed at $A(t) = a$ for all t during the follow-up period.

Thus, the probability of an uncensored individual will experience the event in time t given receiving treatment level a throughout the follow-up period can be written as a product of terms, one per period, describing the conditional probability that the event did not occur since time 0 to $t-1$ but occur in period $(t-1, t]$:

$$\begin{aligned} Pr(t_s = t|A = a) &= h(t, a)(1 - h(t-1, a))(1 - h(t-2, a)) \cdots (1 - h(0, a)) \\ &= h(t, a) \prod_{j=0}^{t-1} (1 - h(j, a)). \end{aligned}$$

Similarly, the probability that a censored individual will experience an event after time t can be written as a product of terms describing the conditional probability that the event did not occur in any observation:

$$\begin{aligned} s(t,a) &= Pr(t_s > t | A = a) \\ &= (1 - h(t,a))(1 - h(t-1,a))(1 - h(t-2,a)) \cdots (1 - h(0,a)) \\ &= \prod_{j=0}^{t} (1 - h(j,a)). \end{aligned} \tag{1}$$

which is also the population survival function.

We use the general propensity score (GPS) to eliminate the biases associated with differences in the covariates.[2] This approach consists of two steps. First, we estimate the conditional expectation of the survival outcome $S(t,a)$ as a function of two scalar variables, the treatment level a and the GPS function $g(\cdot)$. Second, to estimate the dose response function at a particular level of treatment, we average this conditional expectation over the GPS at that treatment value. Specifically, let us follow the definition in Imbens' work[15], where the $g(\cdot)$, the conditional density of the treatment received, is defined as:

$$g(a,x) = f_{A|X}(a|x) \tag{2}$$

and the GPS is $G(a,x) = g(A = a, X = x)$. The GPS has a similar balancing property to that of the standard propensity score. For a stratum with the same value of $g(a,X)$, the probability of receiving treatment a is independent of the value of X:

$$X \perp\!\!\!\perp A = a | g(a,X).$$

This implies that the assignment to treatment is unconfounded given the generalized propensity score:

$$f_A(a|g(a,X), S(t,a)) = f_A(a|g(a,X)). \tag{3}$$

Proof. The LHS of the Equation 3 can be written as:

$$f_A(a|g(a,X), S(t,a)) = \int f_A(a|x, g(a,X), S(t,a))\, dF_X(x|S(t,a), g(a,X)).$$

Under the assumption of weak unconfoundedness, $f_A(a|x, g(a,X), S(t,a)) = f_A(a|x)$, so

$$f_A(a|g(a,X), S(t,a)) = \int g(a,x)\, dF_X(x|S(t,a), g(a,X)) = g(a,X).$$

And the RHS of the Equation 3 can be written as:

$$\begin{aligned} f_A(a|g(a,X)) &= \int f_A(a|x, g(a,X))\, dF_X(x|g(a,X)) \\ &= \int f_A(a|x)\, dF_X(x|g(a,X)) \\ &= \int g(a,x)\, dF_X(x|g(a,X)) = g(a,X). \end{aligned}$$

Therefore, for each a, Equation 3 holds. □

Equation 3 implies that if the assignment to the treatment is weakly unconfounded, given pretreatment variables X, we have:

$$\mathbb{E}[S(t,a)|g(a,X) = g] = \mathbb{E}[S(t)|A = a, G = g)]$$

Proof. Let $f_{S(t,a)|A,g(a,X)}(\cdot|a,g)$ denotes the conditional density of $S(t,a)$ given $A = a$ and $g(a,X) = g$. Then, the Bayes rule and Equation 3 tell us:

$$\begin{aligned} f_{S(t,a)|A,g(a,X)}(s(t)|a,g) &= \frac{f_A(a|S(t,a) = s(t), g(a,X) = g) f_{S(t)|g(a,X)}(s(t)|g)}{f_A(a|g(a,X) = g)} \\ &= f_{S(t,a)|g(a,X)}(s(t)|g) \end{aligned}$$

Hence, we have

$$\begin{aligned}\mathbb{E}[S(t,a)|A=a,G=g] &= \mathbb{E}[S(t,a)|A=a,g(A,X)=g] \\ &= \mathbb{E}[S(t,a)|A=a,g(a,X)=g] \\ &= \mathbb{E}[S(t,a)|g(a,X)=g].\square\end{aligned}$$

Then our causal parameter of interest, the survival conditional average dose response (CADR) function with respect to a subgroup $X = x$ can be denoted as:

$$\psi(t,a,x) = \mathbb{E}\big[\mathbb{E}_{X=x}[S(t,a)|g(a,x)=g]\big], \tag{4}$$

and the population survival average dose response (ADR) function will be:

$$\Psi(t,a) = \mathbb{E}\big[\mathbb{E}[S(t,a)|g(a,X)=g]\big],$$

In the following sections, we may drop the notation t in $\psi(t,a,x)$ and use $\psi(a,x)$ to indicate the entire survival CADR curve over $t \in [1,2,\ldots,\Theta]$ given a and x. In addition, we denote the term $\bar{\psi}(a,x)$ to indicate the mean CADR over time, that is:

$$\bar{\psi}(a,x) = \frac{1}{\Theta}\sum_{t=1}^{\Theta}\psi(t,a,x).$$

2.3 Estimate of a dose response function through recurrent neural works

The previous section describes the approach for identifying and estimating the CADR function for survival outcomes using the GPS approach.Two functions need to be modeled: the conditional expectation $\psi(t,A,X)$ in Equation 4 and the GPS function $g(A,X)$ in Equation 2. We propose to use an ensemble of recurrent neural networks with varying random seeds to model both quantities in opposition to the semi-parametric and parametric model as discussed in previous sections. In particular, as detailed in our previous work on modeling the survival outcome[4], we create the outcome label Y^M for individual i as a matrix over the follow-up period $t \in [1,2,\ldots,\Theta]$:

$$Y^M = \begin{bmatrix} E \\ C \end{bmatrix},$$

where

$E = [e(t=1), e(t=2), \ldots, e(t=t), \ldots, e(t=\Theta)]$
$e(\cdot) = 1$ for $t < t$ if i is censored or having an event at t
$e(\cdot) = 0$ for $t \geq t$;
$C = [c(t=1), c(t=2), \ldots, c(t=t), \ldots, c(t=\Theta)]$
$c(\cdot) = 0$ if i is censored at t;
$c(\cdot) = 0$ and $c(t) = 1$ if i is having an event at t.

The output of our model will be a Θ-dimension vector, $\hat{H}_\Theta$, and each element represents the predicted conditional probability of surviving a time interval, which will be $1 - \hat{h}(t,a)$, for $t \in [0,1,2,\ldots,\Theta]$. Then the estimated survival curve will be given by $\hat{s}(t,a) = \prod_{\tau=0}^{t}(1-\hat{h}(\tau,a))$ for a given treatment level a. We fit an ensemble of simple recurrent neural networks with input from the observed treatment trajectory of each individual and the corresponding GPS. That is $Y^M = f(A,\hat{G})$), where $\hat{G}$ is a plug-in value calculated from the estimated density function $\hat{g}(A,X)$ based on the treatment generating process.

2.3.1 Estimate of an unknown probability density function

The probability density function (PDF) of a data generating process only can be accurately measured by standard statistical distributions. For an unknown data generating process such as the treatment assignment process, we propose to use the recurrent neural network with the basis expansion technique for the univariate response a and pose the conditional dense estimation (CDE) problem as a series of univariate regression problems as detailed in the FlexCode study.[16] Specifically, we will have:

$$g(a_t,x) = f\Big(\sum_j \beta_j(x)\phi_j(a_t)\Big) \tag{5}$$

where a_t is the observed treatment level at time t and x is defined similar as in section 2.2 as the trajectory of historical covariates from $t-u$ to $t-1$ for an individual. $\phi_j(\cdot)$ is an orthonormal basis like a Fourier or wavelet basis for functions of $a_t \in \mathbb{R}$. By the orthogonality property of the basis, the expansion coefficients $\beta_j(x)$ are orthogonal projections of $g(a_t, x)$ onto the basis vectors. We can estimate these coefficients by regressing the transformed response variables $\phi_j(a_t)$ on predictors x for every basis function. Rather than relying on regression methods to estimate the expansion coefficients in Equation 5, we computes the coefficients jointly with a recurrent neural network that minimizes the CDE loss:

$$L(\hat{g}, g) = \int \int \Big(\hat{g}(a_t, x) - g(a_t, x)\Big)^2 \, da_t \, dF_X(x).$$

Since the true $g(a_t, x)$ is unknown, we approximate the loss function by:

$$L(\hat{g}, g) = \frac{1}{n} \sum_{i=1}^{n} \int \hat{g}(a_t, x_i)^2 \, da_t - \frac{2}{n} \sum_{i=1}^{n} \hat{g}(a_i(t), x_i)^2$$

using the trajectories of $\{a, x\}$ observed in the validation dataset. In this study, we use the wavelet basis expansion as discussed in the original FlexCode study.

2.3.2 Choice of hyperparameters and model uncertainty

In this study, we use Bayesian optimization [17] to find the optimal hyperparameters in each model. For the survival outcome model, we tune the following hyperparameters:

- Length of the history window $u \in [1, \Theta]$, where Θ is the maximum follow-up period (This parameter is fixed in the simulation study.);
- Batch size $n \in [\min(128, N), N]$, where N is the sample size;
- Number of dense layer $L \in [1, \Theta]$;
- Number of dense layer units $Lu \in [1, D]$, where D is sample dimension, and;
- Number of gated recurrent units (GRU) [18] $Gu \in [1, \Theta]$.

In addition to the above hyperparameters, the GPS model also tunes the following:

- The number of basis functions $j \in [30, 60]$, the choice of parameter range is by trial and error with various simulation studies that will be discussed in the later sections.

To capture the estimation uncertainty of the proposed model, we create an ensemble of m networks with the same structure but varying random seeds. Hence, the final estimate of the survival outcome will be $Y^M = \frac{1}{m} \sum_{k=1}^{m} f_k(A, \hat{G})$ with each $f_k(\cdot)$ trained with a different random seed yet on the same data. Likewise, the GPS will be modeled by $g(a_t, x) = \frac{1}{m} \sum_{k=1}^{m} f_k\big(\sum_j \beta_j(x)\phi_j(a_t)\big)$. Finally, the estimation of Equation 4 is given by averaging over the combination of $m \times m$ models:

$$\hat{\psi}(a, x) = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{q=1}^{m} f_{s,k}\bigg(a, f_{g,q}\Big(\sum_j \beta_j(x)\phi_j(a)\Big)\bigg),$$

where the subscripts s and g denote the survival and GPS models, respectively, and the calculation of the standard deviation is straightforward. In our simulation and case studies, we choose $m = 25$ as further increasing the m does not change the confidence interval of the estimation by a meaningful amount.

2.4 Integrating causal parameters with automated treatment recommendation

In clinical studies, patients are subject to different levels of risk based on their prognostic features, biomarkers, and treatments received. We generalize this assumption as follows. Let us assume that each treatment level a corresponds to an independent risk function given by the survival CADR function $\psi(a, x)$. Then, we can take the difference of the potential CADR averaged over time to calculate the personal risk ratio of prescribing one treatment option over another. We define this difference of log CADR as the recommender function:

$$r(x, a, a') = log(\hat{\bar{\psi}}(a', x)) - log(\hat{\bar{\psi}}(a, x)) \tag{6}$$

The recommender function can be used to provide personalized treatment recommendations. We first pass a patient through the network once in treatment group a and again in treatment group a' and take the difference. When a patient receives a positive recommendation $r(x, a, a') > 0$, treatment a leads to a higher risk of event than treatment a'. Hence, the patient should be prescribed treatment a' and vice versa.

While this approach is feasible for a set of discrete treatments, for stochastic interventions we face the undesired choice to discretise the continuous treatment options. To avoid this subjective decision, we propose two methods for comparison in this study:

- Random search (RS): This approach is inspired by the random search technique for finding the optimal hyperparameters of a neural network model. Here we pose the optimal treatment option as the hyperparameters in the model. By specifying a searching space

$$\pi \ := \ \{\forall a' \in A\} \quad -\min(A) < a' < \max(A)$$

  bounded by the observed set of treatment levels A. We locate the optimal treatment a^* by calculating the expected recommender value $r(x, a, a^*)$ over randomly selected a' in π and comparing it with the original treatment level a.

  In this study, we limit the search for the optimal treatment at the commencement of the follow-up window. This is because of the fact that by varying the level of initial a' the entire survival curve will respond accordingly during the follow-up window. For fully dynamic treatment decisions, this framework can be easily extended to search a^* at the beginning of each time interval during the follow-up window.

- Reinforcement learning (RL): We performed an evaluation of the actual actions (policy) of clinicians using temporal difference learning (TD-learning) of a state-action value function (Q^π) by observing all prescriptions of treatments in existing records (offline sampling) and computing the average value of each treatment option using the pseudo environment created by the estimated survival CADR function.

  The advantage of TD-learning over policy iteration is that it does not require knowledge of the Markov decision process (MDP) and is model-free, which makes it possible to learn simply from sample trajectories.[19] It was computed iteratively from actual patient episodes of successive state-action pairs using the following updating formula:

$$Q^\pi(a, x) \leftarrow Q^\pi(a, x) + \alpha \cdot (r + \gamma \cdot Q^\pi(a', x') - Q^\pi(a, x)) \tag{7}$$

  With $Q^\pi(a, x)$ being the current {action, state} tuple, $Q^\pi(a', x')$ the next {action, state} tuple, α the learning rate, r the immediate reward and γ, the discount factor. We choose $\gamma = 0.99$ to model the fact that a future reward of higher survival probability is worth as much as the immediate survival probability. In this study, the state refers to the prognostic features and biomarkers observed during the history window. The immediate reward function r is defined as the recommender function $r(x, a, a')$ in Equation 6.

  We learn the optimal policy (which we call the RL policy) for the MDP using policy iteration, which identifies the decisions that maximize the expected survival outcome of patients. Policy iteration started with a random policy that was iteratively evaluated and then improved until converging to an optimal solution. After convergence, the RL policy π^* corresponded to the actions with the highest state action value in each state:

$$\pi^*(x) \leftarrow \arg\max_a Q^{\pi^*}(a, x) \, \forall x$$

  Consistent with the RS approach, we estimate the optimal policy $\pi*$ at the commencement of the follow-up window across patients. Thus, the corresponding value V of a policy π is computed using the one-step Bellman equation for V^π and represented the expected return when starting in x and following π thereafter:

$$V^\pi(x) = \sum_{a'} g(a', x) Q^\pi(a', x)$$

  where $Q^\pi(a', x) = \hat{\psi}(a, x) + r(x, a, a')$.

In both methods, we restrict the set of actions to frequently observed treatment levels (i.e., values fall in the 10^{th} to the 90^{th} percentile of the observed treatment values.) taken by clinicians. As such, the resulting RL policy suggests the best possible treatment among all options chosen (relatively frequently) by clinicians.

3 Study Design

3.1 Simulation studies

Our simulation study assumes the patient's survival probability follows a standard exponential distribution with time-varying risk over time, where the conditional distribution of the hazard rate for each time interval, $h(t, a)$, given X and treatment level a, is

$$h(t,a)|X \sim N(a + \sum_j x_j e^{-a\sum_j x_j}, 1), j \in 1, 2, \ldots, D,$$

and the conditional mean of $h(t, a)$ given X is $a + \sum_j x_j e^{-\sum_j ax_j}$. Suppose also that the marginal distribution of X is unit exponential, thus the marginal mean of $h(t, a)$ is obtained by integrating out the covariate to get

$$\mathbb{E}[h(t,a)] = \mathbb{E}[\mathbb{E}[h(t,a)|X]] = a + \frac{D}{(a+1)^{D+1}}. \tag{8}$$

The derivation is presented in Appendix A. The corresponding survival CADR function following Equation 1 and 4 is

$$\psi(t,a,x) = \mathbb{E}\big[\mathbb{E}_{X=x}[\prod_{j=0}^{t}(1 - h(j,a))|g(a,x) = g]\big].$$

with a and x indicate the trajectory time-varying covariates and treatments, respectively.

We conduct simulations by generating the following variables:

- D continuous covariates $X(0)_1, X(0)_2, \ldots, X(0)_D \sim \mathrm{N}(0, V)$ at time 0, where V is variance of the normal distribution and $D = d$ is the feature dimension. We update their value at time t as $X(t)_d = X(t-1)_d/t^{0.5}$ to construct the time-varying baseline;
- A stochastic exposure: $A \sim Exp(\eta/D\sum_{d=1}^{D} X(t)_d > +(1-\eta)\cdot 0.5)$, where η controls the level of overlapping. When $\eta = 0$, the probability of receiving the treatment is independents of $X(0)$; when $\eta = 1$, the allocation follows the mean of $X(t)$; and when $\eta = 0.5$.
- Censoring probability: $C(t) = \exp(-\frac{\log(t)}{\lambda})$, where $\lambda = 30$;
- Survival probability given by Equation 8: $S(t) = \prod_{j=0}^{t}(1 - h(j, a))$;
- An event indicator generated using *root-finding [3]* at each time t: $E(t) = I(S(t) < U \sim \mathrm{Uniform}(0, 1)))$, with the event time defined by $T = t$ if $E(t) = 1$, otherwise $T = \max(T) + 1$;
- A censoring indicator generated using the *root-finding* technique: $CE(t) = I(C(t) < U \sim \mathrm{Uniform}(0, 1)))$, with the censoring time defined by $C = t$ if $CE(t) = 1$, otherwise $C = \max(T)$;
- Survival outcome given by indicator function: $Y = I(T \leq C)$, and;
- The maximum follow up time is 12 time steps.

A series of experiments were conducted by changing the following parameters: $V \in \{0.5, 1, 2\}, D \in \{4, 8, 20, 40\}$, $\eta \in \{0.1, 0.5, 1\}$, $N \in \{1000, 3000, 5000\}$. In addition to the simulation parameters, we also test the effect of the length of the history window on the accuracy of the estimated survival CADR. In particular, we investigate the length of the history window of size $H = \in \{1, 3, 6\}$. Our default data generation model is created with $V = 0.5, D = 8, \eta = 0.5, N = 3000,$ and $H = 1$. For each scenario, we generate a training sample and a testing sample of the same size with the same parameters but different random seeds. All evaluations are based on testing samples.

3.2 Databases and empirical study design

We performed retrospective empirical analyses using three cohorts:

- The Clinical Practice Research Datalink (CPRD) database. [5] We captured a cohort of 20,270 patients (**AF Age**) with non-valvular atrial fibrillation (AF) receiving either Vitamin K Antagonists (VKAs) or Non-Vitamin K antagonist oral anticoagulants (NOAC) during a follow-up period up to 36 months. To test the validity of the model, we used age as a pseudotreatment and modeled its impact on patient mortality. The null hypothesis

Table 1: Descriptive statistics for databases.

	Count	Mean	SD	0.25	0.75
AF Age					
Unique ID	20,270				
Rows	150,193				
Features	53				
Death (1 = Yes, 0 = No)	2,101 (10.3%)				
Age		75.60	10.645	69.25	83.25
Follow-Up Months (up to 36 month)		12.59	5.880	7.50	19.50
eRI Vesopressor					
Unique ID	6,225				
Rows	98,716				
Features	43				
Death (1 = Yes, 0 = No)	459 (7.4%)				
Vesopressor Dosage (µg/kg/min)		0.29	1.913	0.00	0.14
Follow-Up Hours (up to 48 hours)		35.16	20.354	16.00	52.00
eRI Early Discharge					
Unique ID	3,527				
Rows	22,147				
Features	43				
Discharge (1= Other, 0=Home)	1,310 (37.1%)				
ICU Hours		25.13	13.899	16.00	32.00
Follow-Up Hours (up to 180 hours)		58.87	41.384	24.00	84.00

The outcome and treatment variables of each database are shaded in gray.

for testing is that the recommended age should be the same as the original age of the patients. The summary statistics of the database are presented in Table 1, and a detailed description of the cohort can be found in our previous work. [20]

- The multicenter explantory cohort study utilizing ICU patients in the eICU Research Institute (eRI) database with complete hospitalization between January 1, 2007 and March 31, 2011. Detailed descriptions of the eRI database are provided in the original data report. [6] In the first eRI cohort study, we tackle the problem of the optimal treatment strategies for sepsis in ICU as initially presented by Komorowski [21] from the angle of causal inference by assigning the treatment with the highest expected gain in average survival probability over a period of 48 hours. This cohort (**eRI Vesopressor**) includes 6,225 patients with eRI who experienced sepsis according to the Sepsis-3 definition [22] and having been treated with vasopressors. The interaction between the vasopressor dosage and the mortality of patients up to 48 hours in ICU was modeled.
- In hospitals, Step Down Units (SDUs) provide an intermediate level of care between the Intensive Care Units (ICUs) and the general medical-surgical wards. Because SDUs are less richly staffed than ICUs, they are less costly to operate; however, they also are unable to provide the level of care required by the sickest patients. Using the eRI dataset, we generated the third cohort (**eRI Early Discharge**) with 3,527 patients that have been sent to SDUs to understand whether and when the SDUs should be used. We modeled the relationship between the length of stay in ICU (ICU hours) and the terminal discharge status (i.e., discharged to home or not) of these patients in SDUs up to 180 hours. Compared to the previous study [23] which models the flow dynamics between the ICU and SDU to find the optimal allocation of patients to SDU, our model demonstrates the decision should be based on the optimal timing to send a patient to SDU when the terminal outcome can be potentially improved. We detailed the inclusion criteria of eRI cohorts in Appendix D.

3.3 Model evaluation and benchmark

The performance of DeepSDRF is assessed with simulation studies using the three metrics described below:

Absolute percentage bias (Bias): Defined as the absolute percentage bias in the estimated conditional average treatment effect:

$$\text{Bias}(t) = \frac{1}{N}\sum_i \left|\frac{\hat{\psi}_i(t) - \psi_i(t)}{\psi_i(t)}\right|$$

Coverage ratio: Refers to the percentage of times that the true treatment effect lies within the 95% quantile intervals of the posterior distribution of the estimated individual treatment effect.

$$\text{Coverage}(t) = \frac{1}{N}\sum_i I(|\hat{\psi}_i(t) - \psi_i(t)| < CI_i(t))$$

where I is an indicator function, $I = 1$ if $I(\cdot)$ is true and 0 otherwise. CI is the 95% quantile interval of the estimations.

Root-mean-square error (RMSE): Refers to the expected mean squared error of the estimated individual treatment effect:

$$\text{RMSE}(t) = \frac{1}{N}\sum_i (\hat{\psi}_i(t) - \psi_i(t))^2.$$

To test the performance of DeepSDRF on the estimation of conditional average dose response (CADR), we repeated and averaged the results from 50 iterations of each simulation. The performance of DeepSDRF was benchmarked against a plain recurrent neural network with survival outcome as described in the benchmark algorithms in our previous work [4]: the survival recurrent neural network (SNN), which uses the standardised value of the original features as the regressor instead of using the GPS and the treatment level. We conducted the study using Python 3.9.0 with Tensorflow 2.5.0 [24] (code available at `https://github.com/EliotZhu/DeepSDRF`). To handle the problem of partial observability in the time series data, we use the masking layers as described in previous studies. [25, 4]

4 Simulation Results

4.1 Potential outcome estimation performance of Deep Survival Dose Response Function

Table 2: Simulation study results.

Model	**DeepSDRF**			**SNN**		
	Bias	**Coverage**	**RMSE**	**Bias**	**Coverage**	**RMSE**
By level of dimension (D)						
4	0.062 (0.002,0.121)	0.716 (0.663,0.769)	0.054 (0.053,0.055)	0.104 (0.000,0.212)	0.494 (0.441,0.547)	0.236 (0.233,0.239)
8	0.061 (0.000,0.124)	0.607 (0.561,0.654)	0.043 (0.042,0.044)	0.095 (0.091,0.111)	0.398 (0.362,0.434)	0.225 (0.221,0.229)
20	0.063 (0.009,0.117)	0.614 (0.604,0.623)	0.047 (0.047,0.048)	0.119 (0.105,0.133)	0.336 (0.298,0.374)	0.218 (0.216,0.221)
40	0.068 (0.010,0.125)	0.632 (0.589,0.674)	0.052 (0.052,0.053)	0.129 (0.100,0.158)	0.376 (0.296,0.456)	0.242 (0.239,0.245)
By level of overlap (η)						
Low ($\eta = 0.1$)	0.062 (0.002,0.121)	0.716 (0.562,0.869)	0.054 (0.053,0.055)	0.104 (0.000,0.212)	0.494 (0.241,0.747)	0.236 (0.233,0.239)
Medium ($\eta = 0.5$)	0.061 (0.000,0.124)	0.607 (0.561,0.654)	0.043 (0.042,0.044)	0.095 (0.091,0.111)	0.398 (0.362,0.434)	0.225 (0.221,0.229)
High ($\eta = 1.0$)	0.235 (0.019,0.450)	0.560 (0.360,0.759)	0.077 (0.076,0.078)	0.293 (0.000,0.66)	0.331 (0.104,0.558)	0.294 (0.291,0.297)
By sample size (N) under high level of overlap						
10000	0.277 (0.000,0.624)	0.799 (0.642,0.956)	0.073 (0.072,0.075)	0.331 (0.000,0.751)	0.756 (0.581,0.930)	0.235 (0.232,0.239)
3000	0.235 (0.019,0.450)	0.560 (0.360,0.759)	0.077 (0.076,0.078)	0.293 (0.000,0.660)	0.331 (0.104,0.558)	0.294 (0.291,0.297)
1000	0.236 (0.053,0.419)	0.394 (0.212,0.576)	0.135 (0.134,0.136)	0.279 (0.000,0.641)	0.275 (0.145,0.406)	0.310 (0.308,0.312)
By length of history window (H) in the outcome model						
1	0.061 (0.000,0.124)	0.607 (0.561,0.654)	0.043 (0.042,0.044)	0.095 (0.091,0.111)	0.398 (0.362,0.434)	0.225 (0.221,0.229)
3	0.080 (0.022,0.138)	0.651 (0.501,0.801)	0.081 (0.079,0.082)	0.072 (0.000,0.167)	0.527 (0.392,0.661)	0.161 (0.159,0.163)
6	0.081 (0.000,0.174)	0.764 (0.590,0.938)	0.088 (0.086,0.090)	0.059 (0.000,0.124)	0.680 (0.532,0.827)	0.156 (0.154,0.157)

Estimation standard deviation by sample size (N) under high level of confounding

	DeepSDRF(SD)	**SNN(SD)**
10000	0.044 (0.034,0.054)	0.053 (0.029,0.076)
3000	0.026 (0.015,0.037)	0.067 (0.039,0.095)
1000	0.014 (0.008,0.020)	0.141 (0.088,0.194)

All metrics are averaged over 50 simulations for treatment values in the 15th (included) to the 85th (included) percentile of observed levels under the default parameters except for the sample size (N) and the overlap level (O='High'). The shaded row indicates the default scenario.

We compared the performance of DeepSDRF with the plain recurrent neural network for survival outcomes (SNN) in Table 2. Across four values of sample dimensions, DeepSDRF has significantly outperformed SNN across the three captured metrics. In particular, our proposed causal model has approximately doubled the coverage ratio for the true treatment effect for treatment values in the 15th (included) to the 85th (included) percentile of observed treatment levels.

In Figure 1, we plot the conditional average dose response (CADR) for both models by treatment levels and time and compared them to the truth. In addition to the nominal estimation performance on CADR, DeepSDRF also has better performance in estimating the individual dose response (IDR), and its RMSE is only one-fifth of the SNN.

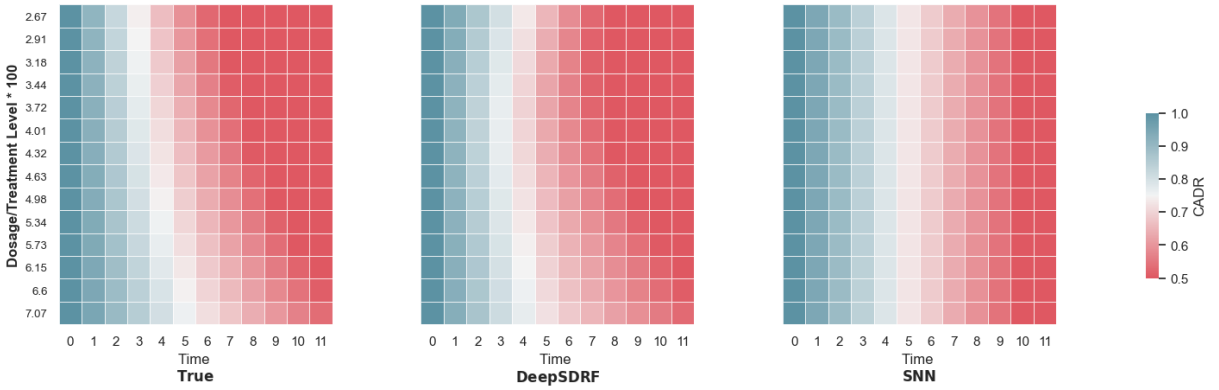

Figure 1: Conditional average dosage response function by benchmark models. Results from DeepSDRF and SNN are from one randomly selected simulation under the default scenario.

An interesting setting for testing causal model performance is to tweak the levels of overlap. In the second section of Table 2, we varied the level of overlap and kept all other parameters under the default setting and found that our estimation has declined accuracy when the level of overlap is high, but has a similar level of accuracy when the level of overlap is low or medium.

To improve the estimation performance under high overlap scenarios, we conducted an analysis with different sample sizes as shown in the third section of Table 2. Increasing the sample size from $3,000$ to $10,000$ has raised the coverage ratio from 0.560 (0.360,0.759) to 0.799 (0.642,0.956), but at the same time, the point estimation Bias for CADR when the sample size is $10,000$ is about 4% worse than the sample size of $3,000$. On the other hand, the RMSE for IDR are at similar levels when the sample size is $3,000$ and above.

The third section of Table 2 examines the scenarios by including history windows from one time step to six time steps in the outcome model of DeepSDRF and SNN. As the GPS model in DeepSDRF has already captured, the historical information, including additional history in the outcome model of DeepSDRF does not improve the performance. In comparison, SNN uses the outcome model to capture the historical information and has improved performance with a longer history window.

The simulations with larger sample size achieved better coverage with higher estimation variance. As shown in the last section of Table 2, we saw the average standard deviation of the CADR estimations across samples is about doubled when the experiments have $10,000$ samples compared to the $3,000$ samples. In contrast to SNN, our estimation variance is increasing with sample size, but the level is lower.

In contrast to binary/discrete treatments, the evaluation of the dose response function for stochastic treatments will face the challenge of estimating the penitential outcomes of rare treatment values. In Table 3, we conducted stress tests on rare values of treatment (i.e., the value falls below the 15th percentile (excluded) or above the 85th percentile (excluded) of the observed treatment values). Compared to Table 2, where the potential outcomes are assessed by more common values, the coverage ratios are about halved when the treatment values are in the lower 15th percentile, and when the treatment values are in the upper 15th percentile, the coverage ratios have declined by about a third. In both tails, DeepSDRF has higher accuracy and lower magnitude of degrading than SNN.

Overall, we found DeepSDRF has stable performance regarding the sample dimension. At the cost of higher estimation variance, we can improve the model performance under high confounding scenarios by increasing the sample size. The separation between the GPS and outcome models makes the DeepSDRF insensitive to the historical windows modeled in the outcome model. The model has more reliable estimations for commonly observed treatment values.

4.2 Stochastic treatment recommendation performance

We applied DeepSDRF to recommend treatment to patients at the commencement of their follow-up. In the first section of Table 4, we record the distribution of the original and recommended treatments from one sample generated under the default scenario. The average recommended treatment by DeepSDRF using random search (RS) is 0.017 (0.0166,0.0169)

Table 3: Stress test

Model	DeepSDRF			SNN		
	Bias	**Coverage**	**RMSE**	**Bias**	**Coverage**	**RMSE**
D	**Treatment percentile <0.15**					
4	0.240 (0.202,0.278)	0.410 (0.315,0.504)	0.115 (0.114,0.116)	0.509 (0.295,0.724)	0.259 (0.133,0.385)	0.334 (0.332,0.336)
8	0.260 (0.161,0.359)	0.480 (0.440,0.520)	0.118 (0.118,0.119)	0.577 (0.324,0.830)	0.175 (0.033,0.317)	0.349 (0.347,0.351)
20	0.333 (0.285,0.381)	0.449 (0.408,0.491)	0.117 (0.116,0.118)	0.459 (0.263,0.655)	0.196 (0.056,0.337)	0.354 (0.352,0.356)
40	0.337 (0.243,0.431)	0.372 (0.300,0.443)	0.133 (0.132,0.134)	0.575 (0.321,0.829)	0.216 (0.080,0.353)	0.361 (0.359,0.364)
	Treatment percentile >0.85					
4	0.078 (0.045,0.111)	0.568 (0.452,0.683)	0.122 (0.121,0.123)	0.312 (0.200,0.424)	0.527 (0.399,0.655)	0.332 (0.329,0.336)
8	0.085 (0.034,0.136)	0.487 (0.443,0.531)	0.153 (0.152,0.154)	0.302 (0.193,0.410)	0.283 (0.138,0.429)	0.315 (0.311,0.319)
20	0.060 (0.039,0.081)	0.442 (0.398,0.487)	0.063 (0.063,0.063)	0.331 (0.212,0.449)	0.327 (0.185,0.469)	0.350 (0.346,0.354)
40	0.063 (0.032,0.094)	0.431 (0.332,0.530)	0.063 (0.062,0.063)	0.278 (0.177,0.378)	0.367 (0.224,0.511)	0.314 (0.310,0.317)

All metrics are averaged over 50 simulations for treatment values in the 15th (included) to the 85th (included) percentile of observed levels under the default parameters except for the sample size (N) and the confounding level (C='High'). The shaded row indicates the default scenario.
The shaded rows indicate the default scenarios.

and using reinforcement learning (RL) is 0.017(0.0167,0.0171). While using the SNN, the recommendation with RS is 0.014(0.0144,0.0145 and with RL is 0.014(0.0141,0.0143). Generally, there is minimal difference between the RS and RL values under both models. Using DeepSDRF, the average treatment value is 0.008 (0.0080,0.0082) higher than the original value.The variance of the recommended treatments is at the same level as the original treatment.

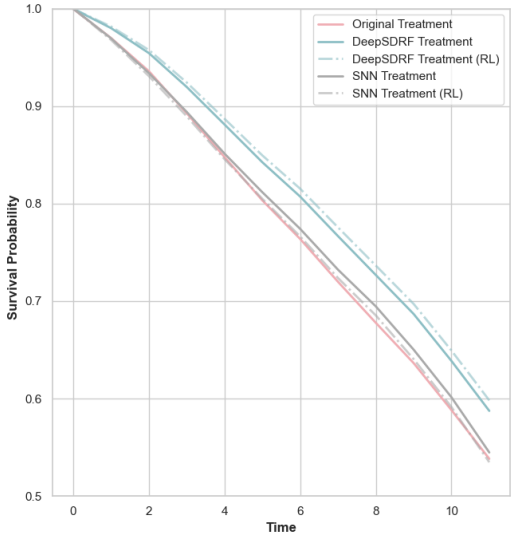

Figure 2: Comparison of survival curves by recommendation schema. The depicted survival curves are averaged over 50 samples generated under the default scenario and following the optimal DeepSDRF or SNN recommendation in each sample.

The second section of Table 4 summarises the average survival probability of DeepSDRF/SNN-recommended patients and the probability estimated using the original treatments. We visualise the corresponding survival curves in Figure 2. Following the recommended treatment plan from DeepSDRF, the average survival probability of the patients is 0.816(0.806,0.826), which is 0.035(0.027,0.043) higher than the original plan, however the survival probability following SNN recommendations has negligible improvement compared to the original. In Table 4, we also observe that the standard deviation of the DeepSDRF outcomes is lower, where the difference between the higher and lower bound of the 95% confidence intervals is 0.004 and the original one is 0.024.

5 Empirical Results

5.1 Empirical validation with pseudo treatments

We apply the random search technique to DeepSDRF in all empirical studies due to its higher efficiency compared to reinforcement learning. We keep the null hypothesis that the DeepSDRF treatments have the same distribution of the

Table 4: Recommendation benchmark

	Recommendation Distribution								
	Default Model (V = 0.5)			**Medium Variance** (V = 1)			**High Variance** (V = 2)		
	Mean (CI)	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**
DeepSDRF-Original	0.008(0.0080,0.0082)	0.007	0.009	0.008(0.008,0.0083)	0.007	0.009	0.008(0.0079,0.0083)	0.006	0.011
Original	0.009(0.0085,0.0088)	0.007	0.010	0.009(0.0085,0.0088)	0.006	0.011	0.009(0.0084,0.0088)	0.005	0.013
DeepSDRF	0.017(0.0166,0.0169)	0.015	0.018	0.017(0.0166,0.0169)	0.014	0.020	0.017(0.0165,0.0170)	0.014	0.022
DeepSDRF (RL)	0.017(0.0167,0.0171)	0.015	0.019	0.017(0.0166,0.0171)	0.014	0.020	0.017(0.0166,0.0172)	0.013	0.023
SNN	0.014(0.0144,0.0145)	0.014	0.015	0.014(0.0144,0.0145)	0.014	0.015	0.014(0.0144,0.0145)	0.014	0.015
SNN (RL)	0.014(0.0141,0.0143)	0.013	0.015	0.014(0.0141,0.0143)	0.013	0.016	0.014(0.0141,0.0143)	0.012	0.017
	Outcome Distribution								
	Overall	**Time 1**		**Time 6**			**Time 12**		
	Mean (CI)	**Mean (CI)**	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**
DeepSDRF-Original	0.035(0.027,0.043)	0.010(0.008,0.012)	-0.003	0.010	0.043(0.035,0.052)	-0.021	0.075	0.049(0.039,0.060)	-0.036	0.113
Original	0.781(0.769,0.793)	0.970(0.967,0.972)	0.969	0.988	0.764(0.755,0.773)	0.723	0.854	0.538(0.528,0.549)	0.455	0.653
DeepSDRF	0.816(0.806,0.826)	0.980(0.978,0.982)	0.979	0.989	0.807(0.801,0.813)	0.787	0.858	0.587(0.580,0.595)	0.545	0.661
DeepSDRF (RL)	0.823(0.813,0.832)	0.982(0.980,0.983)	0.981	0.990	0.815(0.810,0.821)	0.795	0.867	0.598(0.591,0.605)	0.555	0.673
SNN	0.788(0.770,0.806)	0.969(0.961,0.978)	1.000	1.000	0.774(0.755,0.793)	0.697	0.969	0.545(0.529,0.560)	0.406	0.723
SNN (RL)	0.782(0.764,0.800)	0.968(0.959,0.976)	0.998	0.998	0.766(0.748,0.785)	0.690	0.959	0.535(0.519,0.550)	0.398	0.710

Recommendation distribution: The treatment values are depicted from one sample of 3000 patients under the default scenario. Abbreviations: RL, reinforcement learning; CI, 95% confidence interval; 5%, the 5th quantile; 95%, the 95th quantile.

original pseudo-treatments of age (see Figure 3a)[2]. In Table 5a, we see the average difference between the treatment values from the two polices is only 0.11(-0.010,0.230). Figure 3b shows the estimated survival curves following the DeepSDRF recommendations are identical to the survival curves estimated using the original treatments. Figure 3c depicts the dosage response contour of age. When the age increases (ceteris paribus), the mortality of patients declines slightly until 71 years old, after which the mortality increases. Over time, the survival probability drops below 60% after 12 months for most patients.

Table 5: Empirical study results.

Study	**a.Empirical validation**			**b.eRI Vesopressor**			**c. eRI Early Discharge**		
	Mean (CI)	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**	**Mean (CI)**	**5%**	**95%**
Treatment	**Age**			**Vasopressor Dosage, µg/kg/min**			**Length of Stay, hours**		
Original	75.61(75.470,75.763)	69.25	83.25	0.147(0.142,0.152)	0.016	0.196	71.53(70.457,72.614)	48.00	96.00
DeepSDRF	75.72(75.591,75.857)	70.25	84.71	1.059(1.046,1.071)	1.15	1.338	46.77(45.719,47.834)	30.53	51.76
DeepSDRF-Original	0.11(-0.010,0.230)	-1.459	3.041	0.911(0.896,0.926)	0.913	1.271	-24.75(-26.157,-23.361)	-48.78	-2.66
Outcome	**Death**			**Death**			**Death/Discharge to non-home premises**		
Original Outcome	0.915(0.915,0.916)	0.884	0.958	0.939(0.938,0.94)	0.898	0.985	0.574(0.562,0.586)	0.286	0.858
Improved Outcome	0.914(0.913,0.915)	0.888	0.961	0.982(0.981,0.983)	0.988	0.999	0.645(0.634,0.656)	0.401	0.929
Outcome Difference	-0.002(-0.003,-0.0)	-0.004	0.025	0.043(0.042,0.044)	0.001	0.070	0.071(0.063,0.079)	-0.009	0.145

CI, 95% confidence interval; 5%, the 5th quantile; 95%, the 95th quantile.

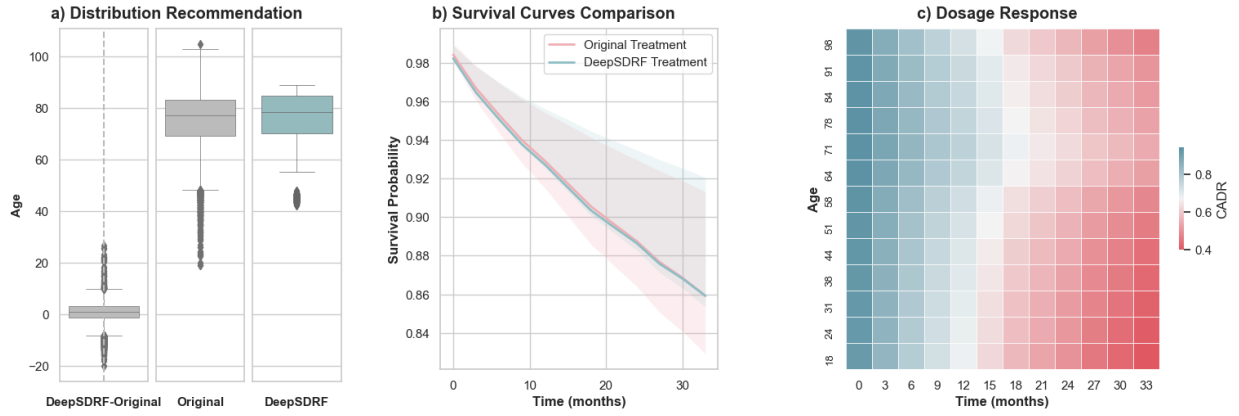

Figure 3: Results for Empirical validation with pseudo treatments.

[2]The Kolmogorov-Smirnov test indicates a p-value of 0.7512.

5.2 Empirical performance with good overlap: optimal vasopressor dosage for sepsis in intensive care

The sepsis case study features a high level of overlap in tertiles of vasopressor dosage : 0.000 to 0.001, 0.001 to 0.075, and above 0.075 µg/kg/min. There is only 5.6% of patients in the comparison of the middle tertile versus others who lack the overlap in values of the GPS close to zero (see Appendix E). Figure 8a shows the distribution of the estimated value of the clinicians' actions and the DeepSDRF policy tested on the eRI cohort. The values of clinicians' policy were averaged at 0.147(0.142,0.152) and the DeepSDRF policies were estimated at and 1.059(1.046,1.071). Figure 8b shows the distribution of patient outcomes according to clinicians' and DeepSDRF policies. On average, the DeepSDRF recommended patients are 4.3% (4.20%,4.40%) more likely to survive during their course of stay in ICU. The response contour in Figure 8c is monotonic, where a higher dosage (ceteris paribus) of the vasopressor is expected to result in better survival outcomes.

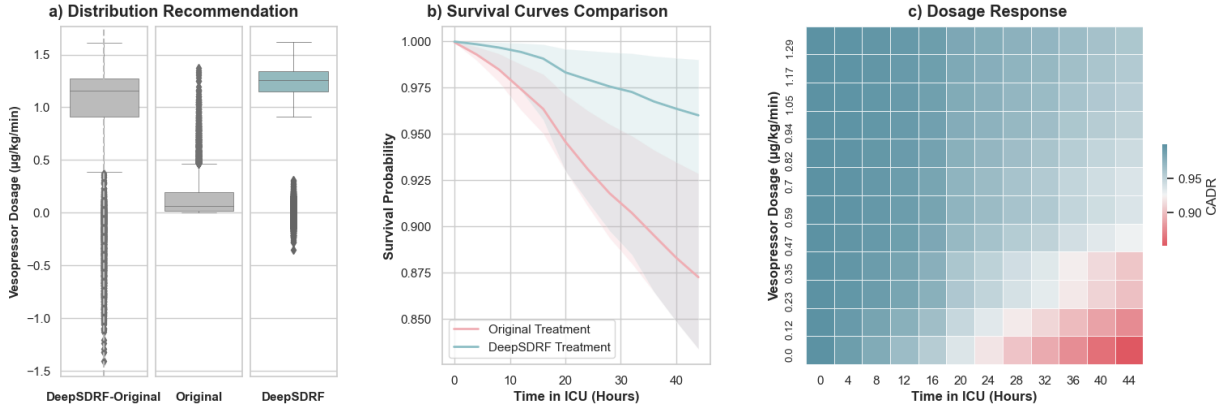

Figure 4: Results for optimal vasopressor dosage for sepsis in intensive care.

5.3 Empirical performance with moderate overlap: the health econometric evaluation of the time to introduce a Step Down Unit

Figure 5a shows the empirical distribution of the ICU length of stay (LoS, hours) before the transfer to the SDUs, while Table 5c describes the recommended LoS. The average hours in ICU following DeepSDRF is 24.75(26.157,23.361) shorter than the observed one. We assessed the estimated GPS in Appendix E using the overlap in tertiles of the LoS: 0.0 to 48.0, 48.1 to 80.0, and above 80.1 hours. Over the follow-up period, it appears that there is 16.2% of patients in the comparison of the middle tertile versus the others, for whom there is a lack of overlap in values of the GPS close to zero. The average probability of being discharged to home has been improved by 7.1% (6.30%,7.91%) following the DeepSDRF suggestions. The relationship between the LoS and the discharge outcome seems to be U-shaped (Figure 5c). However, the CADRs for LoS other than 36 hours find themselves decline below 50% after 96 hours since discharge from ICU, that is, the patients are unlikely to be discharged to home after 4 days in SDU and more than 36 hours in ICU.

6 Discussion

This work introduces the Deep Survival Dose Response Function (DeepSDRF) and illustrates its performance in simulation and empirical examples with characteristics typical of observational clinical evaluations where the sample size is moderate, and it is necessary to control for high-dimension covariates to make a plausible assumption about unconfoundedness. Overall, DeepSDRF achieves nominal performance for estimating the stochastic treatment effects on time-to-event outcomes. To be consistent with other outcomes, we term the dose response surface for survival curves associated with a given value of stochastic exposure as the survival conditional average dose response (CADR) function.

The key contribution of DeepSDRF is the application of the general propensity score (GPS) framework in the context of survival outcomes. The usage of neural network models does not require correctly specifying the parametric model for the GPS or the outcome. We contrast this approach to the parametric implementation of the GPS approach in its original study on the effect of smoking on labor earnings and medical expenditures [26, 8] and the more recent parametric applications on survival outcomes. [27] While we put a particular focus on the improvement achieved by DeepSDRF compared to the plain recurrent neural network for estimating survival outcomes (SNN). The extensive simulations find

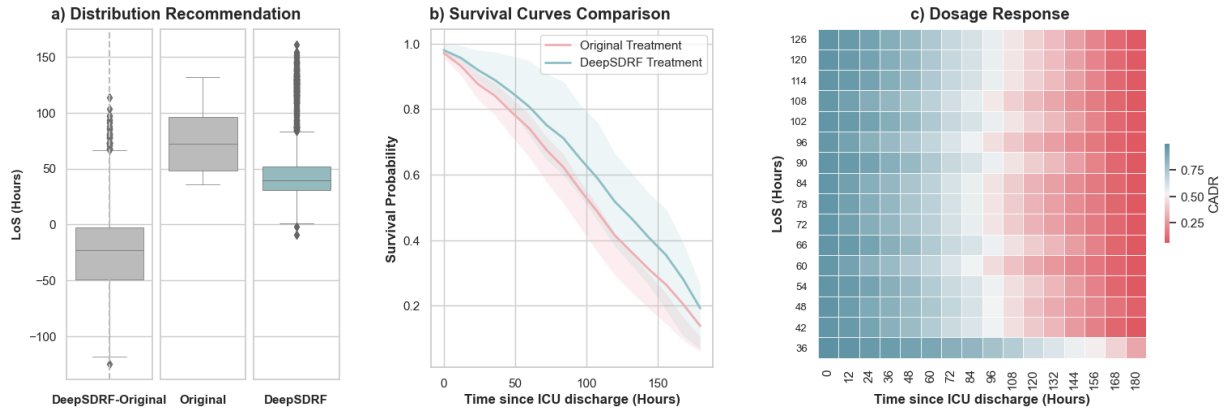

Figure 5: Step Down Unit timing.

the introduction of the GPS into the complex machine learning framework produces superior performance in estimating the CADR function.

Secondly, we propose treatment recommenders using the information derived from the estimated treatment effects. Two approaches have been demonstrated to construct the recommender. The first approach uses the random search technique to efficiently locate the optimal treatment among the potential treatment space. The second approach uses a simplified version of reinforcement learning (or one-step reinforcement learning), which only considers the initial step of the treatment level given the historical information of patients. These two approaches have similar performance in improving patient outcomes, yet the random search is computationally more efficient.

This paper has some limitations. First, the GPS approach reports more conservative confidence intervals than the regression approach. This is expected, as estimators using the propensity score are usually less efficient than estimators based on a correctly specified outcome model.[28]

Second, DeepSDRF relies on the assumption of no unmeasured confounders, specifically in the context of continuous treatment, the weak unconfoundedness assumption. This assumption requires that for any level of treatment, the probability of receiving this level is independent of the potential outcomes, conditional on covariates. In our empirical example for optimal timing for ICU early discharge, this assumption requires that factors that cause delays in discharge, and are also prognostic of the outcomes, are controlled for. We used all potential confounders captured in the eICU datasets without subjective inclusion and exclusion assessment. However, the possibility for unobserved confounding remains, for example, because the covariates are measured at the time of ICU presentation, so subsequent changes in patients' prognosis after discharge, which might cause adverse outcomes during the stay in SDUs, are unmeasured. In the absence of appropriate instrument variables, the effects of unobserved confounders could be examined by employing sensitivity analysis methods in the context of continuous treatment.[29]

Third, in this study, covariate balance following adjustment with the GPS did not improve for all subgroups (indicated by the tertiles of treatment levels that lack support). An alternative loss function for the DeepSDRF could explicitly consider a metric that takes into account the balance achieved. For the binary propensity score, the data adaptive algorithm has been proposed to estimate the GPS based on balance measures such as the Kolmogorov-Smirnoff test. [30] Such approaches still require subjective choices of the appropriate balance measure, the prioritisation of confounders, and for continuous treatment, a method for categorising the treatment variable. Indeed, the most appropriate balance metric remains a topic of ongoing debate.[31]

This work provokes areas of further research. Our data adaptive neural network avoids the misspecification of both the outcome and GPS models but is not capable of improving estimation when the data lacks overlapping among treatment levels. Future simulation studies could examine the sensitivity of the survival CADR function to unmeasured confounding, which is a major cause of the lack of overlapping. Second, the treatment effect estimation and the associated recommendation provided by DeepSDRF is based on a fixed window of historical covariates, future studies can extend this procedure to multiple rolling windows of history given it is possible to change the treatment policy during the estimation period.

Acknowledgment

This work was supported by National Health and Medical Research Council, project grant no. 1125414. Ethics to use UK Clinical Practice Research Datalink data was obtained from ISAC (protocol number 17-093).

A Analytical solution of the treatment effect

The marginal mean of $h(a,t)$ given $j \in [1,2,\ldots,D]$ and $k \in [1,2,\ldots,D]$, where D is the dimension of the covariate space, is obtained as follows:

$$\begin{aligned}\mathbb{E}[\mathbb{E}[h(a,t)|X]] &= \mathbb{E}[a + \sum_j x_j e^{-\sum_j x_j a}] = a + \int_0^\infty \sum_j x_j e^{-\sum_j x_j a} \cdot \prod_j e^{-x_j} dx_j \\ &= a + \frac{D}{(a+1)^{D+1}}\end{aligned} \tag{9}$$

where the integration term can be simplified as:

$$\begin{aligned}&\sum_j \prod_{k:k\neq j} \int_0^\infty \left[e^{-\sum_k x_k(a+1)} \int_0^\infty x_j e^{-x_j(a+1)} dx_j \right] dx_k \\ &= \sum_j \frac{1}{(a+1)^{D+1}} \underbrace{\prod_{k:k\neq j} \int_0^\infty \left[e^{-\sum_k x_k(a+1)} \underbrace{\int_0^\infty x_j e^{-x_j(a+1)} dx_j (a+1)}_{=\Gamma(2)=1} \right] dx_k (a+1)}_{=\Gamma(1)=1}\end{aligned}$$

B Computation of flexible hazard survival function

For each individual i, we define the hazard rate $h(t)$, the probability of experiencing an event in interval $(t-1,t]$, as:

$$h(t) \;:=\; Pr(Y(t)=1 \mid \overline{A}(t,u), \overline{X}(t,u)), \tag{10}$$

where $\bar{A}$ and $\bar{X}$ are the history of treatments and covariates from $t-u$ to $t-1$ with u being the length of the observation history. Thus, the probability that an uncensored individual will experience the event in time t can be written as a product of terms, one per period, describing the conditional probabilities that the event did not occur since time 0 to $t-1$ but occur in period $(t-1,t]$:

$$\begin{aligned}Pr(t_{s,i}=t) &= h(t)(1-h(t-1))(1-h(t-2))\cdots(1-h(0)) \\ &= h(t)\prod_{j=0}^{t-1}(1-h(j)).\end{aligned}$$

Similarly, the probability that a censored individual will experience an event after time t can be written as a product of terms describing the conditional probability that the event did not occur in any observation:

$$\begin{aligned}S(t) &= Pr(t_{s,i}>t) \\ &= (1-h(t))(1-h(t-1))(1-h(t-2))\cdots(1-h(0)) \\ &= \prod_{j=0}^{t}(1-h(j)).\end{aligned} \tag{11}$$

which is the population survival function.

The outcome label Y^M for individual i is defined as a matrix over each time period $t \in [1,2,\ldots,\Theta]$:

$$Y_i^M = \begin{bmatrix} E_i \\ C_i \end{bmatrix},$$

where

$$\begin{aligned} E_i &= (e_i(t=1), e_i(t=2), \ldots, e_i(t=t_i), \ldots, e_i(t=\Theta)) \\ e_i(\cdot) &= 1 \text{ for } t < t_i \text{ if } i \text{ is censored or having an event at } t_i \\ e_i(\cdot) &= 0 \text{ for } t \geq t_i; \\ C_i &= (c_i(t=1), c_i(t=2), \ldots, c_i(t=t_i), \ldots, c_i(t=\Theta)) \\ c_i(\cdot) &= 0 \text{ if } i \text{ is censored at } t_i; \\ c_i(\cdot) &= 0 \text{ and } c_i(t_i) = 1 \text{ if } i \text{ is having an event at } t_i. \end{aligned} \tag{12}$$

The output of our model will be a Θ-dimension vector, $\hat{H}_{i,\Theta}$, and each element represents the predicted conditional probability of surviving a time interval, which will be $\{1 - \hat{h}_i(j)$, for $j \in [0, 1, 2, \ldots, \Theta]\}$. Then the estimated survival curve will be given by $\hat{Y}_i(t) = \prod_{j=0}^{\Theta}(1 - \hat{h}_i(j))$.

C Continuous outcome experiment

We compared the treatment effect estimation performance between general linear model (GLM) and nonparametric neural networks (NN) in the setting of continuous interventions and outcomes. In particular, we examined three types of models to illustrate the improvement achieved by NN, which are NN for both GPS and outcome models; GLM for the GPS model and NN for the outcome model; and GLM for both GPS and outcome models. In this continuous outcome experiment, we simulate the conditional distribution of an outcome Y given X similar to Example 1 in the main study:

$$Y(a)|X \sim N(a + \sum_j x_j e^{-\sum_j ax_j}, 1), j \in 1, 2, \ldots, D,$$

where $a \sim N(0, 1)$ is the value of continuous treatment and D is the dimension of the covariate space. Suppose also that the marginal distributions of X are unit exponential. The marginal mean of $Y(a)$ is obtained by integrating out the covariates to get

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = a + \frac{D}{(a+1)^{D+1}}$$

we present the results at $D = 6$ in Table 6.

Table 6: Experiment results for continuous outcome

	NN		**NN+Linear**		**Linear**	
	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
Correct linear GPS and outcome models						
N=1000	8.838	2.072	4.230	0.750	1.270	0.313
N=3000	5.231	1.058	2.164	0.686	1.227	0.324
N=10000	1.093	0.271	1.300	0.239	1.238	0.304
Misspecified linear GPS and outcome models						
N=1000	8.838	2.072	9.465	3.750	15.668	3.313
N=3000	5.231	1.058	7.172	2.686	15.581	3.327
N=10000	1.093	0.271	5.219	2.239	15.463	3.304

Abbreviations: NN: neural network GPS and outcome models; NN+Linear: linear GPS model and neural network outcome model: Linear: linear GPS and outcome models.

D Descriptive statistics for empirical databases

Table 7: Descriptive statistics for case study 2

	Count	Mean	SD	0.25	0.75
Unique ID	6225				
Rows	98716				
Death (1 = Yes, 0 = No)	459 (7.4%)				
Vesopressor Dosage (µg/kg/min)		0.29	1.913	0.00	0.14
Follow-Up Hours		35.16	20.354	16.00	52.00
Surgery	350 (5.6%)				
Age		64.72	14.074	57.00	75.00
Gender (1 = Male, 0 = Female)	3647 (58.6%)				
Glasgow Coma Scale (GCS)		169.35	16.448	162.60	177.80
Heart Rate (Bp/S)		83.38	28.676	73.15	100.13
Spo2 (%)		89.41	25.472	94.44	98.93
Respiratory Rate (Breaths/Min)		18.24	8.163	15.04	22.85
Non-Invasive BP Systolic (Mmhg)		84.63	51.103	62.25	117.88
Non-Invasive BP Diastolic (Mmhg)		46.17	28.420	28.99	65.25
Non-Invasive BP Mean (Mmhg)		56.16	34.852	0.00	79.25
Temperature (Celsius)		27.61	16.120	0.00	37.20
Shock Index		0.61	0.418	0.00	0.88
Sodium (Mmol/L)		43.46	64.717	0.00	135.00
Potassium (Mmol/L)		1.44	2.016	0.00	3.70
Chloride (Mmol/L)		31.21	48.566	0.00	99.00
Glucose (Mg/Dl)		45.94	81.823	0.00	101.00
Blood Urea Nitrogen (BUN, Mg/Dl)		9.41	19.110	0.00	13.00
Creatinine (Mg/Dl)		0.55	1.206	0.00	0.73
Magnesium (Mg/Dl)		0.42	0.859	0.00	0.00
Calcium (Mg/Dl)		2.23	3.592	0.00	6.90
Total Bilirubin (Mg/Dl)		0.21	1.206	0.00	0.00
AST (SGOT) (Units/L)		62.77	643.250	0.00	0.00
ALT (SGPT) (Units/L)		35.97	319.727	0.00	0.00
Albumin (G/Dl)		0.35	0.928	0.00	0.00
Hgb (G/Dl)		2.94	4.751	0.00	8.00
White Blood Cell Count (K/Mcl)		3.67	7.867	0.00	2.00
Platelets Count (K/Mcl)		42.32	87.025	0.00	34.00
Partial Thromboplastin Time (PTT, Sec)		4.69	16.312	0.00	0.00
Prothrombin Time (PT,Sec)		2.33	7.374	0.00	0.00
International Normalized Ratio (INR)		0.23	0.739	0.00	0.00
Arterial Ph		1.92	3.228	0.00	7.13
Pao2 (Mmhg)		30.49	62.572	0.00	54.00
Paco2 (Mmhg)		10.62	18.797	0.00	25.00
Base Excess (Meq/L)		-0.96	3.644	0.00	0.00
Fio2 (%)		14.13	28.029	0.00	0.00
HCO3 (Mmol/L)		5.27	9.702	0.00	0.00
Lactate (Mmol/L)		0.57	2.069	0.00	0.00
Pre-Admission Fluid Input (Ml)		390.38	2657.466	0.00	0.00
Pre-Admission Fluid Output (Ml)		522.39	2513.307	0.00	100.00
Pre-Admission Balance (Ml)		-132.01	3072.695	0.00	0.00
Fluid Input (Ml/4 Hours)		54.88	316.422	0.00	0.00
Fluid Output (Ml/4 Hours)		32.36	199.217	0.00	0.00
Fluid Balance (Ml/4 Hours)		22.52	328.079	0.00	0.00

Table 8: Descriptive statistics for case study 3

	Count	Mean	SD	0.25	0.75
Unique ID	3527				
Rows	22147				
Discharge (0= Other, 1=Home)	1310 (37.1%)				
ICU Hours		25.13	13.899	16.00	32.00
Follow-Up Hours		58.87	41.384	24.00	84.00
Surgery	205 (5.8%)				
Age		65.03	14.374	57.00	76.00
Gender (1 = Male, 0 = Female)	2010 (57.0%)	0.57	0.495	0.00	1.00
Glasgow Coma Scale (GCS)		169.88	13.855	162.60	178.00
Heart Rate (Bp/S)		67.26	38.184	61.36	92.50
Spo2 (%)		69.67	43.313	0.00	97.80
Respiratory Rate (Breaths/Min)		15.16	9.251	12.00	21.00
Non-Invasive BP Systolic (Mmhg)		81.49	61.090	0.00	128.50
Non-Invasive BP Diastolic (Mmhg)		43.07	32.749	0.00	68.25
Non-Invasive BP Mean (Mmhg)		51.33	40.325	0.00	83.00
Temperature (Celsius)		8.88	15.758	0.00	0.00
Shock Index		0.46	0.364	0.00	0.74
Sodium (Mmol/L)		18.29	46.881	0.00	0.00
Potassium (Mmol/L)		0.58	1.422	0.00	0.00
Chloride (Mmol/L)		13.47	34.741	0.00	0.00
Glucose (Mg/Dl)		17.34	48.259	0.00	0.00
Blood Urea Nitrogen (BUN, Mg/Dl)		3.63	11.954	0.00	0.00
Creatinine (Mg/Dl)		0.17	0.598	0.00	0.00
Magnesium (Mg/Dl)		0.11	0.470	0.00	0.00
Calcium (Mg/Dl)		1.10	2.857	0.00	0.00
Total Bilirubin (Mg/Dl)		0.02	0.230	0.00	0.00
AST (SGOT) (Units/L)		1.53	19.613	0.00	0.00
ALT (SGPT) (Units/L)		2.42	35.095	0.00	0.00
Albumin (G/Dl)		0.09	0.502	0.00	0.00
Hgb (G/Dl)		1.22	3.392	0.00	0.00
White Blood Cell Count (K/Mcl)		1.24	3.786	0.00	0.00
Platelets Count (K/Mcl)		22.56	71.265	0.00	0.00
Partial Thromboplastin Time (PTT, Sec)		0.65	6.665	0.00	0.00
Prothrombin Time (PT,Sec)		0.74	3.894	0.00	0.00
International Normalized Ratio (INR)		0.07	0.359	0.00	0.00
Arterial Ph		0.09	0.806	0.00	0.00
Pao2 (Mmhg)		1.08	11.157	0.00	0.00
Paco2 (Mmhg)		0.59	5.621	0.00	0.00
Base Excess (Meq/L)		0.03	0.746	0.00	0.00
Fio2 (%)		0.30	3.836	0.00	0.00
HCO3 (Mmol/L)		0.28	2.899	0.00	0.00
Lactate (Mmol/L)		0.01	0.133	0.00	0.00
Pre-Admission Fluid Input (Ml)		189.00	753.541	0.00	10.00
Pre-Admission Fluid Output (Ml)		173.14	532.887	0.00	50.00
Pre-Admission Balance (Ml)		15.86	760.122	0.00	0.00
Fluid Input (Ml/4 Hours)		0.38	24.256	0.00	0.00
Fluid Output (Ml/4 Hours)		0.28	15.470	0.00	0.00
Fluid Balance (Ml/4 Hours)		0.11	22.352	0.00	0.00

E Additional results for emprical studies

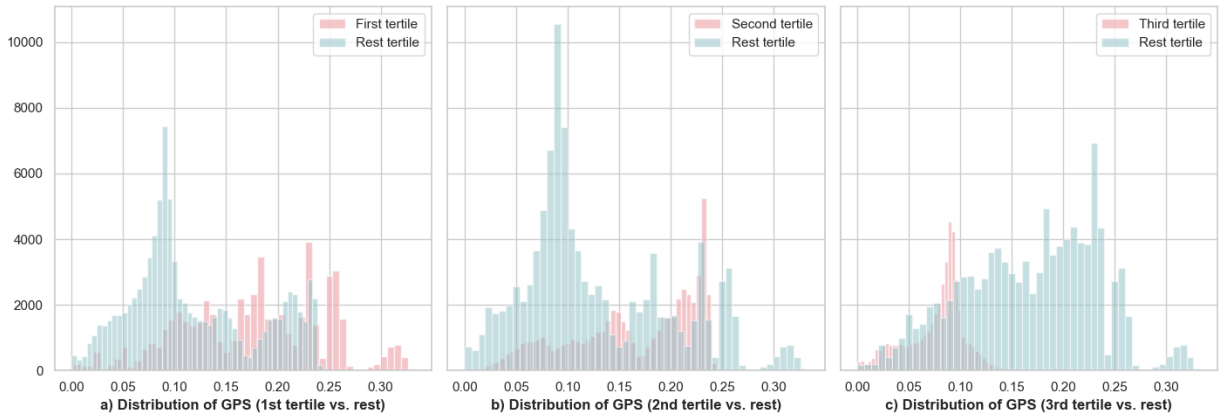

Figure 6: Overlap, based on the GPS estimated at medians of tertiles of the ICU length of stay distributio (AF Age)

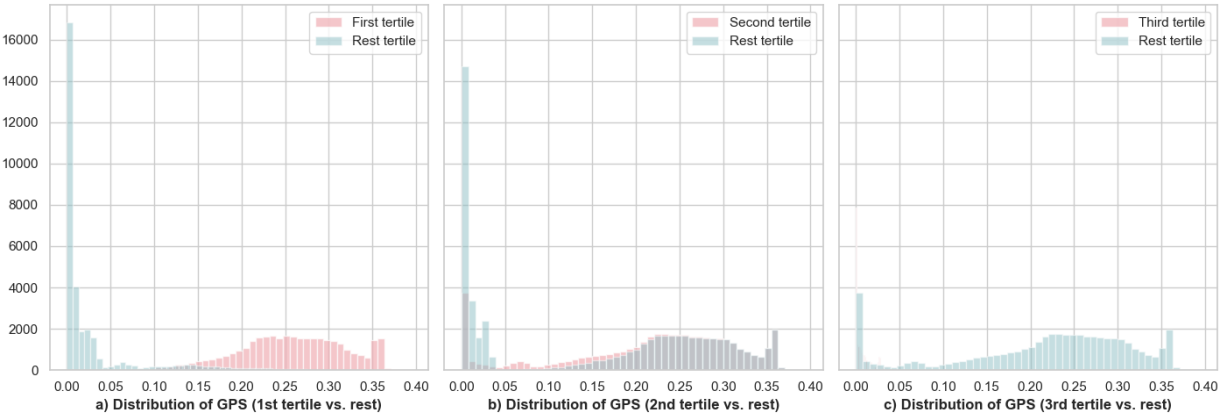

Figure 7: Overlap, based on the GPS estimated at medians of tertiles of the ICU length of stay distribution (eRI Vesopressor)

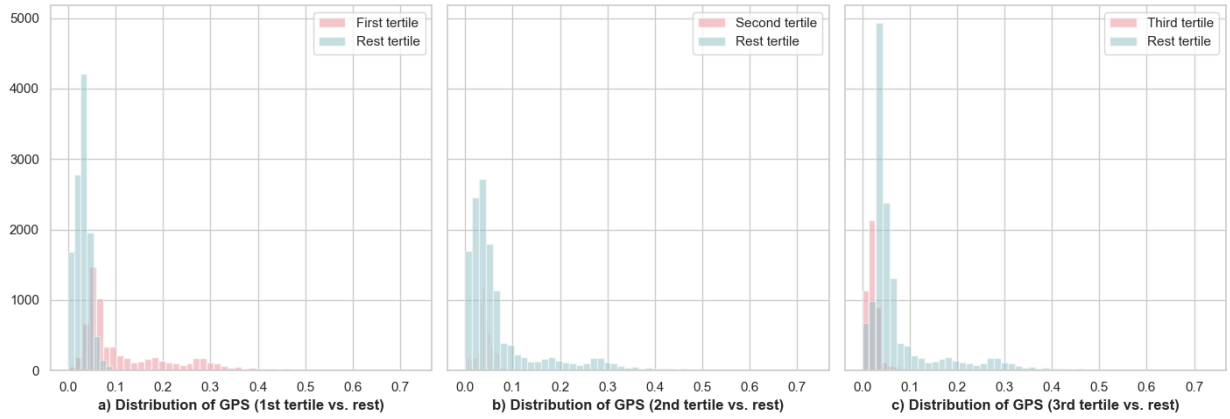

Figure 8: Overlap, based on the GPS estimated at medians of tertiles of the ICU length of stay distribution (eRI Early Discharge)

References

[1] Omer Gottesman, Fredrik Johansson, Joshua Meier, Jack Dent, Donghun Lee, Srivatsan Srinivasan, Linying Zhang, Yi Ding, David Wihl, and Xuefeng Peng. Evaluating reinforcement learning algorithms in observational health settings. *arXiv preprint arXiv:1805.12298*, 2018.

[2] Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

[3] Jie Zhu and Blanca Gallego. Targeted Estimation of Heterogeneous Treatment Effect in Observational Survival Analysis. *Journal of Biomedical Informatics*, page 103474, jun 2020.

[4] Jie Zhu and Blanca Gallego. Casual Inference using Deep Bayesian Dynamic Survival Model (CDS) . *arXiv preprint arXiv:2101.10643*, 2021.

[5] Emily Herrett, Arlene M Gallagher, Krishnan Bhaskaran, Harriet Forbes, Rohini Mathur, Tjeerd Van Staa, and Liam Smeeth. Data resource profile: clinical practice research datalink (cprd). *International journal of epidemiology*, 44(3):827–836, 2015.

[6] Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific data*, 5(1):1–13, 2018.

[7] Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

[8] Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

[9] Antonio F Galvao and Liang Wang. Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment. *Journal of the American Statistical Association*, 110(512):1528–1542, 2015.

[10] Daniel Scharfstein, James M Robins, Wesley Eddings, and Andrea Rotnitzky. Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics*, 57(2):404–413, 2001.

[11] James M Robins, Andrea Rotnitzky, and Daniel O Scharfstein. Statistical models in epidemiology, the environment, and clinical trials. *Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference, ME Halloran and D. Berry (Eds.)*, 95:133, 2000.

[12] Mark J Van der Laan and Sherri Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

[13] Romain Neugebauer and Mark van der Laan. Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434, 2007.

[14] Ori M Stitelman, Victor De Gruttola, and Mark J van der Laan. A general implementation of tmle for longitudinal data applied to causal inference in survival analysis. *The international journal of biostatistics*, 8(1), 2012.

[15] Guido W Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000.

[16] Rafael Izbicki and Ann B Lee. Converting high-dimensional regression to high-dimensional conditional density estimation. *Electronic Journal of Statistics*, 11(2):2800–2831, 2017.

[17] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[18] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[19] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[20] Jie Zhu and Blanca Gallego. Cdsm–casual inference using deep bayesian dynamic survival models. *arXiv preprint arXiv:2101.10643*, 2021.

[21] Matthieu Komorowski, Leo A Celi, Omar Badawi, Anthony C Gordon, and A Aldo Faisal. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature medicine*, 24(11):1716–1720, 2018.

[22] Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810, 2016.

[23] Mor Armony, Carri W Chan, and Bo Zhu. Critical care in hospitals: When to introduce a step down unit. *Product Oper Manag*, 2013.

[24] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[25] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):1–12, 2018.

[26] Michela Bia and Alessandra Mattei. A stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, 8(3):354–373, 2008.

[27] Peter C Austin. Assessing the performance of the generalized propensity score for estimating the effect of quantitative or continuous exposures on survival or time-to-event outcomes. *Statistical methods in medical research*, 28(8):2348–2367, 2019.

[28] Stijn Vansteelandt and Rhian M Daniel. On regression adjustment for the propensity score. *Statistics in medicine*, 33(23):4053–4072, 2014.

[29] Paul R Rosenbaum. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74(1):13–26, 1987.

[30] Brian K Lee, Justin Lessler, and Elizabeth A Stuart. Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3):337–346, 2010.

[31] Jochen Kluve, Hilmar Schneider, Arne Uhlendorff, and Zhong Zhao. Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2):587–617, 2012.