Lifetime Data Analysis (2024) 30:213–236
https://doi.org/10.1007/s10985-023-09606-7

Causal survival analysis under competing risks using longitudinal modified treatment policies

Iván Díaz[1] · **Katherine L. Hoffman**[2] · **Nima S. Hejazi**[3]

Received: 7 February 2022 / Accepted: 17 July 2023 / Published online: 24 August 2023

Abstract

Longitudinal modified treatment policies (LMTP) have been recently developed as a novel method to define and estimate causal parameters that depend on the natural value of treatment. LMTPs represent an important advancement in causal inference for longitudinal studies as they allow the non-parametric definition and estimation of the joint effect of multiple categorical, ordinal, or continuous treatments measured at several time points. We extend the LMTP methodology to problems in which the outcome is a time-to-event variable subject to a competing event that precludes observation of the event of interest. We present identification results and non-parametric locally efficient estimators that use flexible data-adaptive regression techniques to alleviate model misspecification bias, while retaining important asymptotic properties such as $\sqrt{n}$-consistency. We present an application to the estimation of the effect of the time-to-intubation on acute kidney injury amongst COVID-19 hospitalized patients, where death by other causes is taken to be the competing event.

Keywords Modified treatment policies · Competing risks · Targeted minimum loss-based estimation · Double machine learning

✉ Iván Díaz
ivan.diaz@nyu.edu

Katherine L. Hoffman
kh3233@cumc.columbia.edu

Nima S. Hejazi
nhejazi@hsph.harvard.edu

[1] Division of Biostatistics, Department of Population Health, New York University Grossman School of Medicine, New York, NY 10016, USA

[2] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

[3] Department of Biostatistics, T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA

1 Introduction

In survival analysis, a competing event is one whose realization precludes the occurrence of the event of interest. As an example, consider a study on the effect of mechanical ventilation of hospitalized COVID-19 patients on the onset of acute kidney injury, where the death of a patient by extraneous causes may be considered a competing event. Commonly used methods for analyzing time-to-event outcomes under competing events include estimating the cause-specific hazard function (Prentice et al. 1978), the sub-distribution function (also known as cumulative incidence; Fine and Gray 1999), or treating the competing event as a censoring event. Early techniques for the estimation of these parameters were developed in the context of the Cox proportional hazards models (Cox 1972; Andersen and Gill 1982). More recent methods for analyzing survival data include joint modeling (Henderson et al. 2000) and a multitude of other parametric and semi-parametric methods (e.g., Zeger and Liang 1992).

While these methods have allowed much progress in applied research, they have two important limitations: (i) they lack a formal framework that clarifies the conditions required for a causal interpretation of the estimates, and (ii) they typically rely upon parametric modeling assumptions that can induce non-negligible bias in the effect estimates, even when the assumptions required for causal identification hold. Recent developments in causal inference and semi-parametric estimation have yielded important progress in addressing these significant limitations. For example, when the treatment is binary, Young et al. (2020) present a study of the definition and identification of several causal effect parameters in the presence of competing risks, and Benkeser et al. (2018) and Rytgaard and van der Laan (2021) describe semi-parametric–efficient estimators that leverage the flexibility of data-adaptive regression procedures in order to mitigate model misspecification bias while retaining asymptotic properties critical to reliable statistical inference (e.g., $\sqrt{n}$-consistency). Among other insights, Young et al. (2020) clarify that cumulative incidence effects may be interpreted as a total effect of exposure operating through pathways that include the competing event, whereas treating the competing event as a censoring event quantifies direct effects operating independent of the competing event. Furthermore, they clarify that identification when treating the competing event as a censoring event entails considering interventions that eliminate the competing event, and therefore require no unmeasured confounding of the competing event–outcome relation. This assumption is not required to identify the cumulative incidence effect, which is the focus of this work. The reader interested in further issues surrounding the definition and usefulness of these effects in different applications is encouraged to read the original research article (Young et al. 2020).

Under a counterfactual causal framework, defining causal effects requires considering outcomes under hypothetical interventions on the treatment of interest. In the case of binary treatments, causal effects are often defined as differences between the expected outcomes under hypothetical interventions that assign the treatment to all units or to none — that is, the average treatment effect. Causal

effects defined through such deterministic interventions are limited and or lack a basis for scientific interpretation in many applications, including problems where the treatment of interest is numerical rather than categorical, where the treatment is multivariate, where the treatment is a time-to-event-variable (such as time-to-intubation as in our illustrative application), or when treatment assignment is deterministic conditional on potential confounders (cf. positivity/overlap assumptions).

As a solution to the definition of relevant causal effects for these problems, recent work has focused on stochastic interventions and modified treatment policies (Stock 1989; Robins et al. 2004; Díaz and van der Laan 2012; Haneuse and Rotnitzky 2013; Richardson and Robins 2013; Young et al. 2014). Stochastic interventions inquire what would have happened in a hypothetical world where the post-intervention treatment is a random draw from a user-specified distribution conditional on covariates. Examples of stochastic interventions include interventional mediation effects (VanderWeele et al. 2014; Díaz et al. 2020; Hejazi et al. 2022c), as well as incremental propensity score interventions whose identification does not require the positivity assumption (Kennedy 2019; Wen et al. 2021). Modified treatment policies (MTPs) generalize stochastic interventions and allow the post-intervention treatment to also depend on the natural value of treatment (for an overview of such interventions, see, Young et al. 2014). MTPs allow the definition and estimation of parameters defined in terms of natural research questions, such as inquiring what the incidence of wheezing among children with asthma would have been had the concentration of $PM_{2.5}$ particles been 5% lower than they actually were, or what the mortality rate due to opioid overdose would have been had naloxone access laws been implemented a year earlier than they were actually implemented (Rudolph et al. 2021).

Identifying assumptions for causal effects of longitudinal modified treatment policies using the extended g-formula were first articulated by Richardson and Robins (2013) and Young et al. (2014). Recently, Díaz et al. (2021) developed sequential regression and sequentially doubly robust estimators for longitudinal modified treatment policies (LMTP). These estimators are generalizations of estimators developed by Díaz and van der Laan (2018) for MTPs in the single time-point case, and of estimators developed by Luedtke et al. (2017) and Rotnitzky et al. (2017) for static interventions in the multiple time-point case. Here, we extend the LMTP methodology for the estimation of the cumulative incidence under competing events, illustrating how LMTPs can be used to solve problems whose solution has remained elusive in causal inference, such as non-parametric definition and estimation of effects for continuous, multi-valued, time-valued, or multivariate exposures in the context of time-varying exposures. In our motivating example we illustrate estimation of the effect of time-to-intubation on time to acute kidney injury with death by other causes acting as a competing risk.

Our estimation strategy is rooted in recent developments allowing the estimation of causal effects using flexible regression techniques from machine learning (van der Laan and Rose 2011, 2018; Chernozhukov et al. 2017). Central to this theory is the notion of von-Mises expansion (e.g., von Mises 1947; van der Vaart 1998; Robins et al. 2009), which together with sample-splitting techniques (Bickel 1982; Klaassen 1987; Zheng and van der Laan 2011) allows the use of flexible regression methods for the estimation

of nuisance parameters while still allowing the construction of efficient and $\sqrt{n}$-consistent estimators.

2 Defining LMTP effects under competing risks

Assume that at each time point $t = 1, \dots \tau$, we measure $X_t = (D_t, Y_t, L_t, A_t, C_t)$ on each of $i = 1, \dots, n$ study participants, where L_t denotes time-varying covariates, A_t denotes a vector of treatments at time t, C_t denotes an indicator of loss-to-follow-up (equal to one if the unit remains on study at time $t+1$ and equal to zero otherwise), D_t denotes an indicator of the competing event occurring on or before time t, and Y_t denotes an indicator of the event of interest occurring on or before time t. Baseline covariates are included in L_1. At the end of the study, we measure the outcome of interest $Y_{\tau+1}$. We let $D_1 = Y_1 = 0$, meaning that all participants have not experienced the event of interest or the competing event at the beginning of the study. We let $R_t = \mathbb{1}\{D_t = 0, Y_t = 0\}$ denote an indicator that neither the event of interest nor the competing event have occurred by time t. Assume monotone loss-to-follow-up so that $C_t = 0$ implies $C_k = 0$ for all $k > t$, in which case all data become degenerate for $k > t$. In this work, we are interested in a scenario where occurrence of the competing event precludes occurrence of the event of interest. For example, in a study looking to assess the effect of hypertension medications on the likelihood of a stroke, death for reasons other than a stroke is a competing event. In this notation, we have $Y_t = 1$ if $Y_{t-1} = 1$, $D_t = 1$ if $D_{t-1} = 1$, and $Y_t = 0$ if $D_t = 1$ and $Y_{t-1} = 0$, since the competing event precludes occurrence of the event of interest. Let $X_i = (X_{1,i}, \dots, X_{\tau,i})$ denote the observed data for subject i across all time points $t = 1, \dots \tau$. We let $\mathsf{P}f = \int f(x)\, \mathrm{d}\mathsf{P}(x)$ for a given function $f(x)$. We use P_n to denote the empirical distribution of $X_1, \dots X_n$, and assume P is an element of the nonparametric statistical model defined as all continuous densities on X with respect to a dominating measure ν. We let E denote the expectation with respect to P, i.e., $\mathsf{E}\{f(X)\} = \int f(x)\, \mathrm{d}\mathsf{P}(x)$. We also let $||f||^2$ denote the $L_2(\mathsf{P})$ norm $\int f^2(x)\, \mathrm{d}\mathsf{P}(x)$. We use $\bar{W}_t = (W_1, \dots, W_t)$ to denote the past history of a variable W, use $\underline{W}_t = (W_t, \dots, W_\tau)$ to denote the future of a variable, and use $H_t = (\bar{D}_t, \bar{Y}_t, \bar{L}_t, \bar{A}_{t-1})$ to denote the history of the time-varying covariates and the treatment process until just before time t. $\bar{W}$ denotes the entire history $(W_1, \dots, W_\tau)$ of a variable. By convention, variables with an index $t \leq 0$ are defined as the null set, expectations conditioning on a null set are marginal, products of the type $\prod_{t=k}^{k-1} b_t$ and $\prod_{t=0}^{0} b_t$ are equal to one, and sums of the type $\sum_{t=k}^{k-1} b_t$ and $\sum_{t=0}^{0} b_t$ are equal to zero. For vectors u and v, $u \leq v$ denotes point-wise inequalities.

We formalize the definition of the causal effects using a non-parametric structural equation model (Pearl 2000). Specifically, for each time point t, we assume the existence of deterministic functions $f_{D,t}$, $f_{Y,t}$, $f_{L,t}$, $f_{A,t}$, $f_{C,t}$, such that

$$
\begin{aligned}
D_t &= f_{D,t}(C_{t-1}, A_{t-1}, H_{t-1}, U_{D,t}) \\
Y_t &= f_{Y,t}(D_t, C_{t-1}, A_{t-1}, H_{t-1}, U_{Y,t}) \\
L_t &= f_{L,t}(Y_t, D_t, C_{t-1}, A_{t-1}, H_{t-1}, U_{Y,t}) \\
A_t &= f_{A,t}(H_t, U_{A,t}) \\
C_t &= f_{C,t}(A_t, H_t, U_{C,t}).
\end{aligned}
$$

Here $U = (U_{L,t}, U_{A,t}, U_{C,t}, U_{D,t}, U_{Y,t} : t \in \{1, \ldots, \tau + 1\})$ is a vector of exogenous variables. Independence assumptions on these errors necessary for identification will be clarified in Sect. 2.1. Without loss of generality, we assume that variables which are undefined are equal to zero (e.g., $L_t = 0$ if $C_k = 1$ for any $k < t$).

We are concerned with the definition and estimation of the causal effect of an intervention on the treatment process $\bar{A}$ on the event indicator $Y_{\tau+1}$ in a hypothetical world where there is no loss to follow-up (i.e., $\bar{C} = 0$). The interventions on the treatment process are defined in terms of longitudinal modified treatment policies (Díaz and van der Laan 2012; Haneuse and Rotnitzky 2013), which are hypothetical interventions where the treatment is assigned as a new random variable $A_t^{\mathbb{d}}$ (which may depend on the natural value of treatment as explained below) instead of being assigned according to the structural equation model. An intervention that sets the treatments up to time $t-1$ to $\bar{A}_{t-1}^{\mathbb{d}}$ generates a counterfactual variable $A_t(\bar{A}_{t-1}^{\mathbb{d}})$, which is referred to as the *natural value of treatment* (Richardson and Robins 2013; Young et al. 2014), and represents the value of treatment that would have been observed at time t under an intervention carried out up until time $t-1$ but discontinued thereafter. An intervention on all the treatment and censoring variables up to $t = \tau$ generates a counterfactual outcome $Y_{\tau+1}(\bar{A}^{\mathbb{d}})$. Causal effects are defined as contrasts between the counterfactual probability $\mathsf{P}[Y_{\tau+1}(\bar{A}^{\mathbb{d}}) = 1]$ of a failure event by the end of the study under interventions implied by different interventions $\mathbb{d}$, as defined below.

The causal effects we define are characterized by a user-given function $\mathbb{d}(a_t, h_t, \varepsilon_t)$ that maps a given treatment value a_t (i.e., the "natural value") and a history h_t into a new treatment value. The function $\mathbb{d}$ is also allowed to depend on a randomizer ε_t. This definition generalizes static interventions (if $\mathbb{d}(a_t, h_t, \varepsilon_t)$ is a constant), dynamic interventions (if $\mathbb{d}(a_t, h_t, \varepsilon_t)$ depends on h_t but not on a_t nor ε_t), and stochastic interventions (if $\mathbb{d}(a_t, h_t, \varepsilon_t)$ depends on h_t and ε_t but not on a_t). In what follows we will assume that the randomizer is drawn independently across units and independently of U, and whose distribution does not depend on P. For fixed values $\bar{a}_t$, and $\bar{l}_t$, we recursively define $a_t^{\mathbb{d}} = \mathbb{d}(a_t, h_t^{\mathbb{d}}, \varepsilon_t)$, where $h_t^{\mathbb{d}} = (\bar{a}_{t-1}^{\mathbb{d}}, \bar{l}_t)$. The intervention is undefined if $h_t^{\mathbb{d}}$ is such that $d_s^{\mathbb{d}} = 1$ or $y_s^{\mathbb{d}} = 1$ for any $s \leq t$. The LMTPs that we study are thus defined as $A_t^{\mathbb{d}} = \mathbb{d}(A_t(\bar{A}_{t-1}^{\mathbb{d}}), H_t(\bar{A}_{t-1}^{\mathbb{d}}), \varepsilon_t)$ for a user-given function $\mathbb{d}$. In what follows, we let $\mathsf{g}_{A,t}(a_t \mid h_t)$ denote the density or probability mass function of A_t conditional on $C_{t-1} = R_t = 1$ and $H_t = h_t$, and we let $\mathsf{g}_{C,t}(a_t, h_t)$ denote $\mathsf{P}(C_t = 1 \mid C_{t-1} = R_t = 1, A_t = a_t, H_t = h_t)$. Furthermore, we use $\mathsf{g}_t^{\mathbb{d}}(a_t \mid h_t)$ to denote the density or probability mass function of $\mathbb{d}(A_t, H_t, \varepsilon_t)$ conditional on $C_{t-1} = R_t = 1$ and $H_t = h_t$.

To ground the ideas and illustrate the usefulness of this framework, we now consider several examples of regimes d that yield interesting and scientifically informative causal contrasts. More examples appear in Robins et al. (2004); Young et al. (2014); Díaz et al. (2021).

Example 1 (Additive shift LMTP) Let A_t denote a vector of numerical treatments, such as a drug dose or energy expenditure through physical activity. Assume that A_t is supported as $\mathsf{P}(A_t \leq u_t(h_t) \mid H_t = h_t) = 1$ for some u_t. Then, for a user-given value δ, we let

$$\mathrm{d}(a_t, h_t) = \begin{cases} a_t + \delta & \text{if } a_t + \delta \leq u_t(h_t) \\ a_t & \text{if } a_t + \delta > u_t(h_t). \end{cases} \tag{1}$$

We define $A_t^{\mathrm{d}} = \mathrm{d}(A_t(\bar{A}_{t-1}^{\mathrm{d}}), H_t(\bar{A}_{t-1}^{\mathrm{d}}))$. This intervention has been discussed by Díaz and van der Laan (2012), Díaz and van der Laan (2018), and Haneuse and Rotnitzky (2013) in the context of single time-point treatments; by Díaz and Hejazi (2020) and Hejazi et al. (2022c) in the context of causal mediation analysis; and by Hejazi et al. (2020) in the context of studies with two-phase sampling designs. This intervention considers hypothetical worlds in which the natural treatment at time t is increased by a user-given value δ, whenever such an increase is feasible for a unit with history $H_t(A_{t-1}^{\mathrm{d}})$. Specifically, note that if the treatment is supported in the real numbers, the MTP could simply be defined as $\mathrm{d}(a_t, h_t) = a_t + \delta$. However, when the treatment is supported as $\mathsf{P}(A_t \leq u_t(h_t) \mid H_t = h_t) = 1$, the intervention $\mathrm{d}(a_t, h_t) = a_t + \delta$ will not be feasible for units for which $a_t + \delta > u_t(h_t)$. We therefore set $\mathrm{d}(a_t, h_t) = a_t$ for those units. Note that setting $\mathrm{d}(a_t, h_t) = u_t(h_t)$ could also be an interesting option for these units. However, this option would lead to a parameter that violates Condition 5 below, which is necessary to achieve $\sqrt{n}$–consistent estimation.

Example 2 (Multiplicative shift LMTP) Let A_t denote a vector of numerical treatments such as pollutant concentrations for various pollutants. To define this intervention, assume that A_t is supported as $\mathsf{P}(A_t \geq l_t(h_t) \mid H_t = h_t) = 1$ for some l_t. Then, for a user-given value $0 < \delta < 1$, we let

$$\mathrm{d}(a_t, h_t) = \begin{cases} a_t \times \delta & \text{if } a_t \times \delta \geq l_t(h_t) \\ a_t & \text{if } a_t \times \delta < l_t(h_t). \end{cases} \tag{2}$$

The intervention implied by this function considers hypothetical worlds where the natural treatment at time t is increased by a user-given factor δ, whenever such increase is feasible for a unit with history $H_t(A_{t-1}^{\mathrm{d}})$. This intervention seems useful, for example, to solve current problems in environmental epidemiology related to the effect of environmental mixtures (Gibson et al. 2019).

Example 3 (Incremental propensity score interventions based on the odds and risk ratio) Kennedy (2019) proposed an intervention tailored to binary treatments in which the post-intervention treatment is a draw from a user-given distribution

$\mathsf{g}_t^{\mathrm{d}}(a_t \mid h_t)$, where this distribution is chosen such that the odds ratio comparing the likelihood of treatment under intervention versus the actual treatment is a user-given value δ. Specifically, consider $\varepsilon_t \sim U(0,1)$, and define $\mathrm{d}(h_t, \varepsilon_t) = \mathbb{1}\{\varepsilon_t < \mathsf{g}_t^{\mathrm{d}}(1 \mid h_t)\}$ where we define $\mathsf{g}_t^{\mathrm{d}}(1 \mid h_t) = \delta \mathsf{g}_t(1 \mid h_t)/\{\delta \mathsf{g}_t(1 \mid h_t) + \mathsf{g}_t(0 \mid h_t)\}$. Identifying the effect of this intervention does not require the positivity assumption. Thus, incremental propensity score interventions are very useful in settings where violations of the positivity assumption preclude identification of static effects such as the average treatment effect.

Recently, Wen et al. (2021) proposed a similar intervention that instead uses the risk ratio to quantify the likelihood of treatment under the intervention. Their intervention is defined in terms of a random draw from a distribution $\mathsf{g}_t^{\mathrm{d}}(a_t \mid h_t) = a_t \delta \mathsf{g}_t(1 \mid h_t) + (1 - a_t)(1 - \delta \mathsf{g}_t(1 \mid h_t))$ (compared to Wen et al. (2021) we have flipped the values 0 and 1 for ease of exposition). The following LMTP yields the same identifying functional as that of Wen et al. (2021):

$$\mathrm{d}(a_t, \varepsilon_t) = \begin{cases} a_t & \text{if } \varepsilon_t < \delta \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where ε_t is a random draw from the uniform distribution on (0, 1). As before, we make use of the definition $A_t^{\mathrm{d}} = \mathrm{d}(A_t(\bar{A}_{t-1}^{\mathrm{d}}), H_t(\bar{A}_{t-1}^{\mathrm{d}}))$. Note that the probability of treatment under the intervention is equal to $\mathsf{g}_\delta(1 \mid h_t) = \delta \times \mathsf{g}(1 \mid h_t)$; thus, giving a risk-ratio interpretation to δ. Importantly, a unit with zero probability of treatment conditional on their history also has zero probability of treatment under the intervention. This weakens the positivity assumptions required for identification and estimation, as we will discuss below.

Example 4 (Dynamic treatment initiation strategies with grace period) Let A_t denote a binary treatment variable. Consider an intervention of the form "for a user-given grace period m, if a condition for treatment initiation is met at or before $t - m$ then start treatment at t; otherwise, do not intervene." Letting L_t' denote an indicator that the condition has been met, this intervention can be operationalized as an LMTP as follows:

$$\mathrm{d}(a_t, h_t) = \begin{cases} 1 & \text{if } l'_{t-m} = 1 \\ a_t & \text{otherwise.} \end{cases} \tag{4}$$

This type of intervention has been considered in the context of treatment initiation for HIV (van der Laan et al. 2005; Cain et al. 2010), where the condition L_t' is that CD4+ cell count drops below a pre-specified value. In this type of application, realistic treatment policies may involve a delay of treatment initiation for logistical reasons. Accordingly, allowing for a grace period quantifies an effect that is arguably of more practical relevance.

Remark 1 Wen et al. (2021) study stochastic interventions defined as a random draw from a density that can be expressed as a mixture between a known density $\mathsf{f}_t(a_t \mid h_t)$

and the density of the observed data $\mathsf{g}_t(a_t \mid h_t)$. The functional identifying the mean counterfactual outcome under this stochastic intervention is the same as the functional identifying the counterfactual outcome under an LMTP defined by

$$\mathrm{d}(a_t, h_t, \varepsilon_t) = \mathbb{1}\{\varepsilon_t \leq c_t(h_t)\}q_t + \mathbb{1}\{\varepsilon_t > c_t(h_t)\}a_t,$$

where ε_t is a random draw from a uniform distribution in [0, 1], $0 < c_t(h_t) < 1$ is a mixture constant, and q_t represents a draw from $\mathsf{f}_t(a_t \mid h_t)$. Efficient estimators for this type of LMTP have been previously proposed by Díaz et al. (2021). Wen et al. (2021) discuss the stochastic interventions version of these effects, but incorrectly claim that the estimators they develop are not particular cases of the methodology of Díaz et al. (2021).

2.1 Identification of the effect of LMTPs under competing risks

The following assumptions will be sufficient to prove identification:

Assumption 1 (Supported treatment intervention) If $(a_t, h_t) \in \text{supp}\{A_t, H_t\}$ then $(\mathrm{d}(a_t, h_t, \varepsilon_t), h_t) \in \text{supp}\{A_t, H_t\}$ for $t \in \{1, \ldots, \tau\}$ and all values of ε_t.

Assumption 2 (Positivity of loss-to-follow-up mechanism) $\mathsf{P}\{\mathsf{g}_{C,t}(A_t, H_t) > 0\} = 1$ for $t \in \{1, \ldots, \tau\}$.

Assumption 3 (Strong sequential randomization of treatment mechanism) $U_{A,t} \perp\!\!\!\perp (\underline{U}_{D,t+1}, \underline{U}_{Y,t+1}, \underline{U}_{L,t+1}, \underline{U}_{A,t+1}) \mid (H_t, R_t = C_t = 1)$ for all $t \in \{1, \ldots, \tau\}$.

Assumption 4 (Sequential randomization of loss-to-follow-up mechanism) $U_{C,t} \perp\!\!\!\perp (\underline{U}_{D,t+1}, \underline{U}_{Y,t+1}, \underline{U}_{L,t+1}, \underline{U}_{A,t+1}) \mid (A_t, H_t, R_t = 1)$ for all $t \in \{1, \ldots, \tau\}$.

Assumptions 1 and 2 are required so that the interventions considered are supported in the data. In words, Assumption 1 requires that if there is an individual with treatment level a_t and covariate values h_t who is event-free and uncensored at time t, there must also be an individual with treatment level $\mathrm{d}(a_t, h_t, \varepsilon_t)$ and covariate values h_t who is event-free and uncensored at time t. Assumption 1 would be violated if the intervention d is infeasible at time t for a participant with covariate history h_t. As an example of this, consider our illustrative application, letting A_t be an indicator of having been intubated at time t, and letting H_t denote a high-dimensional vector containing all the information collected for a patient up until time t, including information on blood oxygen saturation. Consider a modified treatment policy that would delay intubation by a number of days δ for a large value δ. Note that intubation is a salvage therapy that is almost always used when a patient reaches a low oxygen

saturation point. Therefore, Assumption 1 may be violated for this LMTP since a patient with very low blood oxygen saturation who is intubated at time t has a very low chance of not being intubated in the vicinity of time t.

Assumption 2 states that for every possible value of the covariates h_t that has positive probability of occurring, there are uncensored individuals. Assumption 3 is required to identify the effect of an LMTP. It states that, among event-free uncensored individuals, H_t contains all the common causes of treatment at time t and future events, treatments, and covariates. Assumption 4 is required to identify the effect of a static intervention that prevents loss to follow-up. It states that, among event-free uncensored individuals, H_t contains all the common causes of censoring at time t as well as future events and covariates.

Proposition 1 (Identification of the effect of LMTPs with time-to-event outcomes subject to competing events) *Set* $\mathsf{q}_{\tau+1} = Y_{\tau+1}$ *and* $R_{\tau+1} = 1$. *For* $t = \tau, \ldots, 1$, *recursively define*

$$\mathsf{q}_t : (a_t, h_t) \mapsto \mathsf{E}\left[R_{t+1} \times \mathsf{q}_{t+1}(A_{t+1}^{\mathbb{d}}, H_{t+1}) \mid C_t = R_t = 1, A_t = a_t, H_t = h_t\right], \quad (5)$$

and define $\theta = \mathsf{E}[\mathsf{q}_1(A_1^{\mathbb{d}}, L_1)]$. *Under Assumptions* 1-4, $\mathsf{P}[Y_{\tau+1}(\bar{A}^{\mathbb{d}}) = 1]$ *is identified as* θ.

The identification result above, proved in the supplementary materials, extends the identification result of Young et al. (2020) to LMTPs, and the identification result of Díaz et al. (2021) to the case of competing events. This result allows researchers to estimate from data the counterfactual probability $\mathsf{P}[Y_{\tau+1}(\bar{A}^{\mathbb{d}}) = 1]$, interpreted as the outcome rate in a hypothetical world where the modified treatment policy $\mathbb{d}$ was implemented.

Next, we present efficiency theory and efficient estimators for θ, based on an extension of our prior work (Díaz et al. 2021). These estimators allow the use of slow-converging data-adaptive regression techniques to obtain estimators of θ that converge at parametric rates. The general approach involves finding a von-Mises-type approximation (von Mises 1947; van der Vaart 1998; Robins et al. 2009; Bickel et al. 1997) for the parameter $\mathsf{q}_{L,t}$, which can be intuitively understood to be a first-order expansion with second-order error (remainder) terms. As the errors in the expansion are second-order, the resulting estimator of θ will be $\sqrt{n}$-consistent and asymptotically normal as long as the second-order error terms converge to zero at rate $\sqrt{n}$. This would be satisfied, for example, if all the regression functions used for estimation converge at rate $n^{1/4}$ (see Sect. 3). The reader interested in this general theory is encouraged to consult the works of Buckley and James (1979); Rubin and van der Laan (2007); Robins (2000); Robins et al. (1994); van der Laan and Robins (2003); Bang and Robins (2005); van der Laan and Rubin (2006); van der Laan and Rose (2011, 2018); Luedtke et al. (2017); Rotnitzky et al. (2017).

3 Efficient estimation via sequentially doubly robust regression

As noted by Díaz et al. (2021), non-parametric $\sqrt{n}$-consistent estimation is not possible for arbitrary functions $\mathfrak{d}$ when the treatment is continuous. Thus, we impose the following assumption, which is a generalization of an assumption first used by Haneuse and Rotnitzky (2013). In what follows, we will assume that the function $\mathfrak{d}$ does not depend on the distribution of the observed data P.

Assumption 5 (Piecewise smooth invertibility) Assume that the cardinality of A_t is p. For each h_t, assume the support of A_t conditional on $H_t = h_t$ may be partitioned into p-dimensional subintervals $\mathcal{I}_{t,j}(h_t) : j = 1, \ldots, J_t(h_t)$ such that $\mathfrak{d}(a_t, h_t, \varepsilon_t)$ is equal to some $\mathfrak{d}_j(a_t, h_t, \varepsilon_t)$ in $\mathcal{I}_{t,j}(h_t)$ and $\mathfrak{d}_j(\cdot, h_t, \varepsilon_t)$ has inverse function $\mathfrak{b}_j(\cdot, h_t, \varepsilon_t)$ with Jacobian $\mathfrak{b}_j'(\cdot, h_t, \varepsilon_t)$ with respect to a_t.

In what follows, we assume that either (i) A_t is a discrete random variable for all t, or (ii) A_t is a continuous random variable and the modified treatment policy $\mathfrak{d}$ satisfies Condition 5. Next, for continuous A_t we define

$$\mathsf{g}^{\mathfrak{d}}_{A,t}(a_t \mid h_t) = \sum_{j=1}^{J_t(h_t)} \int \mathbb{1}_{t,j}\{\mathfrak{b}_j(a_t, h_t, \varepsilon_t), h_t\} \times \mathsf{g}_{A,t}\{\mathfrak{b}_j(a_t, h_t, \varepsilon_t) \mid h_t\} \times \left|\det\{\mathfrak{b}_j'(a_t, h_t, \varepsilon_t)\}\right| \mathrm{d}\mathsf{P}(\varepsilon_t), \tag{6}$$

where $\mathbb{1}_{t,j}\{u, h_t\} = 1$ if $u \in \mathcal{I}_{t,j}(h_t)$ and $\mathbb{1}_{t,j}\{u, h_t\} = 0$ otherwise. Under Condition 5, application of the formula for the density of a transformation shows that the p.d.f. of $A_t^{\mathfrak{d}}$ conditional on the history h_t is $\mathsf{g}^{\mathfrak{d}}_{A,t}(a_t \mid h_t)$. In the case of Example 2, the post-intervention p.d.f. becomes

$$\mathsf{g}^{\mathfrak{d}}_{A,t}(a_t \mid h_t) = \mathsf{g}_{A,t}(a_t \mid h_t)\mathbb{1}\{a_t\delta < l_t(h_t)\} + \delta^{-1}\mathsf{g}_{A,t}(a_t\delta^{-1} \mid h_t)\mathbb{1}\{a_t \geq l_t(h_t)\},$$

which shows that piecewise smoothness is sufficient to handle interventions such as (2) which are not smooth in the full range of the treatment. Condition 5 and expression (6) ensure that we can use the change of variable formula when computing integrals over $\mathcal{A}_t$ for continuous treatments. This is useful for studying properties of the parameter and estimators we propose.

For discrete treatments, we let

$$\mathsf{g}^{\mathfrak{d}}_{A,t}(a_t \mid h_t) = \sum_{s_t \in \mathcal{A}_t} \int \mathbb{1}\{\mathfrak{d}(s_t, h_t, \varepsilon_t) = a_t\}\mathsf{g}_{A,t}(s_t \mid h_t)\, \mathrm{d}\mathsf{P}(\varepsilon_t). \tag{7}$$

Evaluated using the definition of the incremental propensity score intervention based on the risk ratio given in Example 3, this expression yields

$$\mathsf{g}^{\mathfrak{d}}_{A,t}(a_t \mid h_t) = a_t\delta\mathsf{g}_{A,t}(1 \mid h_t) + (1 - a_t)[1 - \delta\mathsf{g}_{A,t}(1 \mid h_t)],$$

which clarifies that the risk ratio for receiving the treatment under the intervention versus the observed data is equal to δ. In what follows, it will be useful to define the parameter

$$\mathsf{w}_t(c_t, a_t, h_t) = \frac{\mathsf{g}^{\mathbb{d}}_{A,t}(a_t \mid h_t)}{\mathsf{g}_{A,t}(a_t \mid h_t)} \frac{c_t \times r_t}{\mathsf{g}_{C,t}(a_t, h_t)},$$

and the functions

$$\varphi_t : x \mapsto \sum_{s=t}^{\tau} \left(\prod_{k=t}^{s} \mathsf{w}_k(c_k, a_k, h_k) \right) \{ r_{s+1} \mathsf{q}_{s+1}(a^{\mathbb{d}}_{s+1}, h_{s+1}) - \mathsf{q}_s(a_s, h_s) \} + \mathsf{q}_t(a^{\mathbb{d}}_t, h_t) \tag{8}$$

for $t = \tau, \ldots, 1$ which map $x = (d_t, y_t, l_t, a_t, c_t : t \in \{1, \ldots, \tau\})$ to the real line. When necessary, we use the notation $\varphi_t(x;\eta)$ or $\varphi_t(x;\underline{\eta}_t)$ to highlight the dependence of φ_t on $\underline{\eta}_t = (\mathsf{w}_t, \mathsf{q}_t, \ldots, \mathsf{w}_\tau, \mathsf{q}_\tau)$. We also use η to denote $(\mathsf{w}_1, \mathsf{q}_1, \ldots, \mathsf{w}_\tau, \mathsf{q}_\tau)$, and define $\varphi_{\tau+1}(X;\eta) = Y$.

Estimation of θ will proceed by regression of the transformation φ_{t+1} sequentially from $t = \tau$ to $0 = 1$ to obtain estimates q_t. We say that φ_{t+1} is a sequentially doubly robust unbiased transformation for q_t due to the following proposition:

Proposition 2 (Sequentially doubly robust transformation) *Let η' be such that either $\mathsf{q}'_s = \mathsf{q}_s$ or $\mathsf{w}'_s = \mathsf{w}_s$ for all $s > t$. Then, in the event $R_t = 1$, we have*

$$\mathsf{E}\left[R_{t+1} \times \varphi_{t+1}(X;\eta') \mid C_t = R_t = 1, A_t = a_t, H_t = h_t\right] = \mathsf{q}_t(a_t, h_t).$$

This proposition is a straightforward application of the von-Mises expansion given in Lemma S1 in the supplementary materials. Another consequence of Lemma S1 is that the efficient influence function for estimating $\theta = \mathsf{E}[\mathsf{q}_1(A^{\mathbb{d}}, L_1)]$ in the non-parametric model is given by $\varphi_1(X) - \theta$ (see Díaz et al. 2021). Proposition 2 motivates the construction of the sequential regression estimator by iteratively regressing an estimate of the data transformation $\varphi_{t+1}(X;\eta)$ on (A_t, H_t) among individuals with $C_{t-1} = R_t = 1$, starting at $\varphi_{\tau+1}(X;\eta) = Y$.

Remark 2 (Double robustness of incremental propensity score interventions) Consider the odds ratio incremental propensity score interventions of Example 3, where we let $\varepsilon_t \sim U(0, 1)$, and define $\mathbb{d}(h_t, \varepsilon_t) = \mathbb{1}\{\varepsilon_t < \mathsf{g}^{\mathbb{d}}_t(1 \mid h_t)\}$. The result in Proposition 2 also holds for this intervention, but it does not allow for doubly robust estimation. To see why, notice that q_t itelse depends on $\{\mathsf{g}_s : s > t\}$ (or equivalently, on $\{\mathsf{w}_s : s > t\}$) through $\mathbb{d}$. Therefore, if $\mathsf{q}'_s = \mathsf{q}_s$ and $\mathsf{w}'_s \neq \mathsf{w}_s$ for all $s > t$, we get $\mathsf{E}\left[R_{t+1} \times \varphi_{t+1}(X;\eta') \mid C_t = R_t = 1, A_t = a_t, H_t = h_t\right] = \mathsf{q}_t(a_t, h_t)$, where q_t is evaluated at the misspecified $\{\mathsf{w}'_s : s > t\}$. This is aligned with the findings of Kennedy (2019). In contrast, the risk ratio incremental propensity score intervention does allow for doubly robust estimation, since $\mathbb{d}_t$ does not depend on w_t.

In order to avoid imposing entropy conditions on the initial estimators, we use sample splitting and cross-fitting (Klaassen 1987; Zheng and van der Laan

2011; Chernozhukov et al. 2018). Let $\mathcal{V}_1, \ldots, \mathcal{V}_J$ denote a random partition of the index set $\{1, \ldots, n\}$ into J prediction sets of approximately the same size. That is, $\mathcal{V}_j \subset \{1, \ldots, n\}$; $\bigcup_{j=1}^{J} \mathcal{V}_j = \{1, \ldots, n\}$; and $\mathcal{V}_j \cap \mathcal{V}_{j'} = \emptyset$. In addition, for each j, the associated training sample is given by $\mathcal{T}_j = \{1, \ldots, n\} \setminus \mathcal{V}_j$. We let $\hat{\eta}_j$ denote the estimator of η obtained by training the corresponding prediction algorithm using only data in the sample $\mathcal{T}_j$. Further, we let $j(i)$ denote the index of the validation set containing unit i. For preliminary estimates $\hat{\mathsf{w}}_{1,j(i)}, \ldots, \hat{\mathsf{w}}_{\tau,j(i)}$, the estimator is defined as follows:

Step 1 Initialize $\varphi_{\tau+1}(X_i; \check{\underline{\eta}}_{\tau,j(i)}) = Y_i$ for $i = 1, \ldots, n$.
Step 2 For $t = \tau, \ldots, 1$:

- Compute the pseudo-outcome $\check{Y}_{t+1,i} = R_{t+1} \times \varphi_{t+1}(X_i; \check{\underline{\eta}}_{t,j(i)})$ for all $i = 1, \ldots, n$.
- For $j = 1, \ldots, J$:
  - Regress $\check{Y}_{t+1,i}$ on $(A_{t,i}, H_{t,i})$ using any regression technique and using only data points $i \in \mathcal{T}_j$.
  - Let $\check{\mathsf{q}}_{t,j}$ denote the output of the above regression, update $\check{\underline{\eta}}_{t,j} = (\hat{\mathsf{w}}_{t,j}, \check{\mathsf{q}}_{t,j}, \ldots, \hat{\mathsf{w}}_{\tau,j}, \check{\mathsf{q}}_{\tau,j})$, and iterate.

Step 3 Define the sequential regression estimator as

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \varphi_1(X_i; \check{\eta}_{j(i)}).$$

Implementation of the above estimator requires an algorithm for estimation of the parameters w_t for all t. This parameter may be estimated component-wise as follows. First, the censoring mechanism $\mathsf{g}_{C,t}$ may be estimated through any regression or classification procedure by regressing the indicator C_t on data (A_t, H_t) among individuals with $R_t = C_t = 1$. Thus, it only remains to estimate the density ratio $\mathsf{g}_t^{\mathbb{d}}(a_t \mid h_t)/\mathsf{g}_t(a_t \mid h_t)$. An option to estimate this density ratio is to first obtain an estimate of the density $\mathsf{g}_t(a_t \mid h_t)$, and then to obtain an estimator of $\mathsf{g}_t^{\mathbb{d}}(a_t \mid h_t)$ by applying formula (6) or (7). This approach is problematic in the case of continuous or multivariate treatments due both to the curse of dimensionality and to the fact that flexible and computationally efficient estimation of conditional densities is an under-developed area of machine learning (relative to regression), though some suitable techniques do exist (e.g., Díaz and van der Laan 2011; Hejazi et al. 2022b, a). This approach can also be problematic for discrete treatments of high cardinality. Since the conditional densities themselves are unnecessary for estimation, and direct estimation of the ratio $\mathsf{g}_t^{\mathbb{d}}(a_t \mid h_t)/\mathsf{g}_t(a_t \mid h_t)$ suffices, we propose to use Díaz et al. (2021)'s approach to direct estimation of density ratios. Briefly, this approach starts by constructing an augmented artificial dataset of size $2n$ in which each observation is duplicated. For each observation, one duplicate gets assigned the post-intervention treatment value $\mathbb{d}(A_t, H_t)$ and the other gets the actual observed value A_t. Then, the task of density ratio estimation reduces to the task of classifying which record

belongs to the intervened-upon unit, based on the treatment and the history H_t. Specifically, Díaz et al. (2021) show that the odds of this classification problem is equal to the density ratio. Importantly, data-splitting procedures such as cross-validation or cross-fitting must be performed such that the two records from the same participants are always kept in the same fold.

The above estimation algorithm requires the implementation of several regression procedures. Many algorithms from the machine learning literature are available for such regression problems, including those based on ensembles of regression trees, splines, etc. As stated in the asymptotic normality result below, it is important that the regression procedures used approximate as well as possible the relevant components of the true data-generating mechanism. Increasing the predictive performance of these regression fits can entail estimator selection from a large list of candidate estimators. If the sample size is large enough, a principled solution to this problem is the use of Super Learning (van der Laan et al. 2006, 2007), a form of model stacking (Wolpert 1992; Breiman 1996) that constructs an ensemble model as a convex combination of candidate algorithms from a user-supplied regression library, where the weights in the combination are computed based on cross-validation and are such that they minimize the holdout prediction error of the resulting estimator.

The asymptotic distribution of θ is Gaussian with variance equal to the efficiency bound, under conditions. To state the conditions, we first define the data-dependent parameter

$$\check{\mathsf{q}}_t^{\dagger}(a_t, h_t) = \mathsf{E}\big[R_{t+1}\varphi_{t+1}(X; \underline{\check{\eta}}_t) \mid C_t = R_t = 1, A_t = a_t, H_t = h_t\big],$$

where the outer expectation is only with respect to the distribution P of X (i.e., $\check{\eta}$ is fixed). The following theorem is a consequence of Theorem 4 of Díaz et al. (2021):

Theorem 1 (Consistency and weak convergence of SDR estimator) *We have*

(i) *If, for each time t, either $||\hat{\mathsf{w}}_t - \mathsf{w}_t|| = o_{\mathsf{P}}(1)$ or $||\check{\mathsf{q}}_t - \check{\mathsf{q}}_t^{\dagger}|| = o_{\mathsf{P}}(1)$, then $\hat{\theta} = \theta + o_{\mathsf{P}}(1)$.*

(ii) *If $\sum_{t=1}^{\tau} ||\hat{\mathsf{w}}_t - \mathsf{w}_t|| \, ||\check{\mathsf{q}}_t - \check{\mathsf{q}}_t^{\dagger}|| = o_{\mathsf{P}}(n^{-1/2})$ and $\mathsf{P}\{\mathsf{w}_t(A_t, H_t) < c\} = \mathsf{P}\{\hat{\mathsf{w}}_t(A_t, H_t) < c\} = 1$ for some $c < \infty$, then*

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_1(X_i; \eta) + o_P(1),$$

implying that $\sqrt{n}(\hat{\theta} - \theta) \rightsquigarrow N(0, \sigma^2)$, where $\sigma^2 = \mathsf{Var}\{\varphi_1(X; \eta)\}$ is the nonparametric efficiency bound.

Note that this means that $\hat{\theta}$ is *sequentially doubly robust* in the sense that it is consistent if at least one of the two nuisance parameters is consistently estimated at each time point. This property is also known in the literature as 2^τ*-multiple robustness*. We prefer the term *sequentially doubly robust* for two reasons. First, contrasted to the double robustness property which requires one out of two nuisance estimators to be consistent, the term 2^τ-multiple robustness erroneously conveys the message that consistency would follow from one out of 2^τ nuisance estimators being consistent. Second, retaining the term "double" in this description conveys a fundamental property of the estimation procedure: it is based on an expansion (given in Lemma S1 in the supplementary material) with second-order remainder terms. In this nomenclature, a triply robust estimating procedure would be one based on an expansion with third-order remainder terms, and so forth.[1]

A drawback of the above estimation approach is that there is no guarantee that the parameter will remain within the bounds of the parameter space. That is, it is possible, in principle, that the sequential regression estimator will provide time-to-event probabilities outside of the closed unit interval [0, 1]. In the supplementary materials, we present a targeted minimum loss estimator (TMLE) capable of producing estimates that always lie within the unit interval. However, boundedness of the TMLE comes with a robustness tradeoff. In particular, the robustness properties of the TMLE are inferior to those of the sequentially doubly robust estimator. Specifically, consistency of the TMLE requires that either $\|\hat{\mathsf{w}}_t - \mathsf{w}_t\| = o_{\mathsf{P}}(1)$ or $\|\hat{\mathsf{q}}_t - \mathsf{q}_t\| = o_{\mathsf{P}}(1)$, for an estimator $\hat{\mathsf{q}}_t$ constructed based on sequential regression using (5). As a result, consistent estimation of q_t in the TMLE procedure requires consistent estimation of $\mathsf{q}_s : s > t$. By contrast, the sequentially doubly robust estimator provides consistent estimation of q_t when either q_s or w_s is consistently estimated for all $s > t$. Both Luedtke et al. (2017) and Díaz et al. (2021) provide further in-depth discussion on this topic. Furthermore, the conditions required for $\sqrt{n}$-consistency for this TMLE algorithm may be stronger than those required for the SDR estimator. In the supplementary materials, we provide an argument that the second-order term associated to the TMLE is bounded above by the term

$$\sum_{t=1}^{\tau} \left(\|\hat{\mathsf{w}}_t - \mathsf{w}_t\| \sum_{s=t}^{\tau} \|\hat{\mathsf{q}}_s - \mathsf{q}_s^\star\| \right),$$

where $\mathsf{q}_t^\star$ is the data-dependent parameter defined as

$$\mathsf{q}_t^\star(a_t, h_t) = \mathsf{E}\left[R_{t+1}\hat{\mathsf{q}}_{t+1}(a_{t+1}, h_{t+1}) \mid C_t = R_t = 1, A_t = a_t, H_t = h_t\right].$$

In comparison to the assumption in statement (ii) of the above theorem, convergence of the second-order term associated to the TMLE seems to require stronger assumptions that would also require convergence of all the cross-product terms $\|\hat{\mathsf{w}}_t - \mathsf{w}_t\|\|\hat{\mathsf{q}}_s - \mathsf{q}_s^\star\|$ for $s \geq t$.

[1] Some of these points were brought to our attention by Edward H. Kennedy in a conversation on Twitter (see this thread, https://twitter.com/edwardhkennedy/status/1446952129218949128?s=20 &t=8o_-20szw0PPfqocYiw7Dg).

Neither the SDR nor the TMLE guarantee that a survival function estimated at several points in time will result in monotonic decreasing estimates. In order to leverage the desirable robustness properties of the SDR estimator while endowing it with the properties of boundedness and monotonicity, we leverage a set of techniques proposed by Westling et al. (2020). Under their approach, out-of-bounds estimates are truncated to remain within bounds, and the possibly non-monotonic estimate of the survival curve, alongside any confidence region limits, is projected onto the space of monotonic functions by way of isotonic regression. Importantly, Westling et al. (2020) show that the resulting projected estimator retains the same asymptotic properties as the original estimator. In our case, this implies that the projected estimator retains the properties of asymptotic normality and consistency under the assumptions of Theorem 1. Specifically, simultaneous confidence bands centered around the isotonic projection constructed using the multivariate distribution implied by Theorem 1 will have correct asymptotic coverage under the assumptions of the theorem.

Under the conditions of the above theorem, a Wald-type procedure, in which the standard error may be estimated based on the empirical variance of $\varphi_1(X_i;\check{\eta}_{j(i)})$, yields asymptotically valid confidence intervals and hypothesis tests. Furthermore, the asymptotic linearity of Theorem 1 implies that estimates of survival functions for multiple time points asymptotically converge to a multivariate normal distribution, which can be used to construct simultaneous confidence intervals using the method described, e.g., by Hothorn et al. (2008).

4 Illustrative application

4.1 Background on motivating study and data

To illustrate the proposed methods, we aim to answer the following question: *what is the effect of invasive mechanical ventilation (IMV) on acute kidney injury (AKI) among COVID-19 patients?* This question has been recently identified by an expert panel on lung–kidney interactions in critically ill patients (Joannidis et al. 2020) as an area in need of further research. AKI is a common condition in the ICU, complicating about 30% of ICU admissions, and causing increased risk of in-hospital mortality and long-term morbidity and mortality (Kes and Jukić 2010). AKI is often viewed as a tolerable consequence of interventions to support other failing organs, such as IMV (Husain-Syed et al. 2016).

To answer this question, we will use data on approximately 3,300 patients hospitalized with COVID-19 at the New York Presbyterian Cornell, Queens, or Lower Manhattan hospitals. Our analysis focuses on patients hospitalized between March 3rd and May 15th, 2020, who did not have a previous history of Chronic Kidney Disease (CKD). This choice of timeframe is rooted in there being a lack of treatment guidelines in the initial months of the COVID-19 pandemic, which resulted in a large variability in provider practice regarding mechanical ventilation. This clinical heterogeneity in the deployment of IMV makes the problem particularly well-suited to assessment via the causal effects of LMTPs. The analytical dataset was

built from two distinct databases. Demographics, comorbidity, intubation, death, and discharge data were abstracted from electronic health records by trained medical professionals into a secure RedCAP database (Goyal et al. 2020). These data were supplemented with the Weill Cornell Critical carE Database for Advanced Research (WC-CEDAR), a comprehensive data repository housing laboratory, procedure, diagnosis, medication, microbiology, and flowsheet data documented as part of standard care (Schenck et al. 2021). This dataset is unique in that it contains a large sample of patients treated in the same hospital system within a short period of time, while still exhibiting extensive treatment variability.

The most frequent reason for morbidity and mortality among patients with COVID-19 is pneumonia leading to acute respiratory distress syndrome (ARDS). Respiratory support via devices such as nasal cannulae, face masks, or endotracheal tubes (i.e., intubation and subsequent mechanical ventilation) is a primary treatment for ARDS (Hasan et al. 2020). For severe ARDS patients, an important decision must be made regarding the optimal timing of intubation based on clinical factors such as oxygen saturation, dyspnea, respiratory rate, chest radiograph, and, occasionally, external resource considerations such as ventilator or provider availability (Tobin 2020; Thomson and Calligaro 2021). At the outset of the COVID-19 pandemic, multiple international guidelines recommended early intubation in an effort to protect healthcare workers from infection and to avoid complications stemming from "crash" intubations (Papoutsi et al. 2021). Yet, as the pandemic progressed and multiple reports documented high mortality for mechanically ventilated patients, guidance changed to delaying intubation with the rationale that IMV increases risk of secondary infections, ventilator-induced lung injury, damage to other organs, and death (Bavishi et al. 2021; Tobin 2006). The lungs and kidneys are physiologically closely tied, and it has been hypothesized that mechanical ventilation may cause AKI in COVID-19 patients via oxygen toxicity and capillary endothelial damage leading to inflammation, hypotension, and sepsis (Durdevic et al. 2020).

Despite these hypotheses, a causal link between intubation and AKI in patients with COVID-19 has not been definitively established. Furthermore, such a link is difficult to study in a randomized fashion due to the clinically subjective nature of intubation assignment — that is, it is impossible to pick an oxygen saturation breakpoint to safely intubate across all patients and hospital settings (Tobin et al. 2020; Perkins et al. 2020).

While the original dataset cannot be made publicly available due to patient privacy considerations, we have created a synthetic dataset that mimics the general structure of the real clinical dataset, and we have made this synthetic dataset available publicly. The `R` code and the synthetic dataset are available in the GitHub repository at https://github.com/kathoffman/lida-comprisks.

4.2 Time-varying variables and modified treatment policy

For each patient, study time begins on the day of hospitalization. The treatment of interest is daily maximum respiratory support, which can take three levels: no supplemental oxygen, supplemental oxygen not including IMV, and IMV. Our goal is

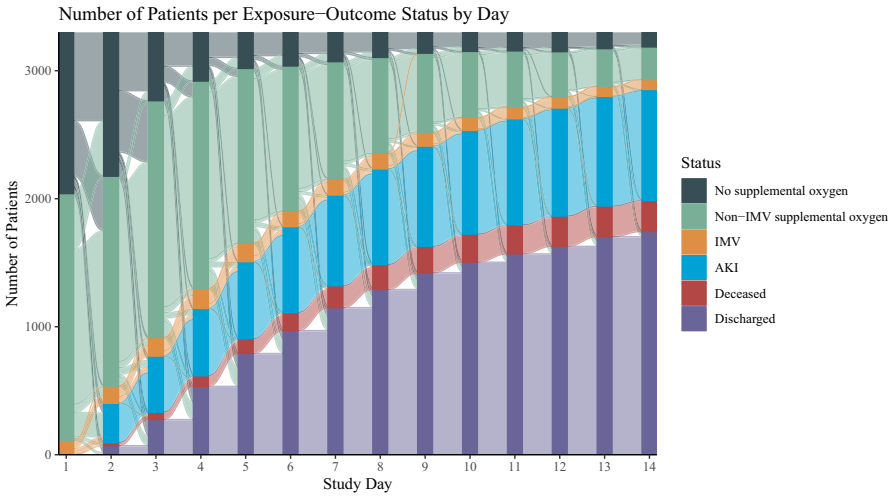

Fig. 1 A flow diagram, or "alluvial plot", showing the movement of patients between exposure and outcome statuses. The *x* axis shows study day and the *y* axis shows counts of patients. Colors indicate exposure status for patients still in the study, or outcome status for patients who have already experienced the outcome, competing risk, or loss-to-follow-up. Darker-colored bars indicate the number of patients in each exposure or outcome at that time point, while lighter-colored bars visualize patients moving between time points. The exposure contains three levels: no supplemental oxygen, supplemental oxygen not including IMV, and IMV. The outcome of interest is AKI, and death is a competing risk. Patients who are discharged are right-censored. There is a one-day lag (i.e. no exposure status shown for the last study day) so that all data used in the analysis can be visualized on a single plot

to estimate the overall causal effect on daily AKI rates of an intervention that would delay intubation for IMV by a single day among 449 patients who received IMV for more than one day, and a prevention of IMV for patients who received IMV for exactly one day. To simplify the presentation, in what follows we refer to this intervention as a "one-day delay in intubation." Death by other causes is treated as a competing risk for AKI development. In notation, this LMTP may be expressed as follows: Let $A_t \in \{0, 1, 2\}$, where 0 indicates no supplemental oxygen at end of day t, 1 indicates supplemental oxygen not including IMV, and 2 indicates IMV. Consider the following intervention:

$$\mathrm{d}_t(a_t, h_t) = \begin{cases} 1 & \text{if } a_t = 2 \text{ and } a_s \leq 1 \text{ for all } s < t, \\ a_t & \text{otherwise.} \end{cases} \tag{9}$$

Under this intervention, a patient who is first intubated at day t would receive non-invasive oxygen support at day t. Otherwise, the patient would receive the oxygen support actually (i.e., "naturally") observed at day t. This intervention assesses what would have happened in a hypothetical world where intubation procedures were more conservative (operationalized as a one-day delay) than they were at the beginning of the pandemic in New York City's surge conditions, when limited information and treatments were available for COVID-19. Figure 1 shows the observed supplemental oxygen levels and observed AKI and death rates across days 1–14.

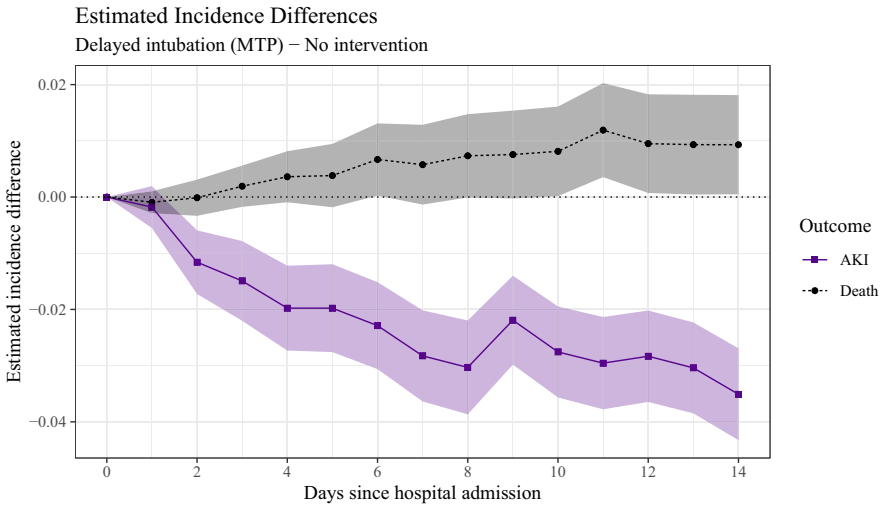

Fig. 2 Difference between the two treatment policies for two outcomes: AKI and mortality. 95% simultaneous confidence bands cover the sets of point estimates

Baseline confounders include age, sex, race, ethnicity, body mass index (BMI), hospital location, home oxygen status, and comorbidities (e.g., hypertension, history of stroke, Diabetes Mellitus, Coronary Artery Disease, active cancer, cirrhosis, asthma, Chronic Obstructive Pulmonary Disease, Interstitial Lung Disease, HIV infection, and immunosuppression), and are included in L_1. Time-dependent confounders include vital signs (e.g., highest and lowest respiratory rate, oxygen saturation, temperature, heart rate, and blood pressure), laboratory results (e.g., C-Reactive Protein, BUN Creatinine Ratio, Activated Partial Thromboplastin time, Creatinine, lymphocytes, neutrophils, bilirubin, platelets, D-dimer, glucose, arterial partial pressure of oxygen, and arterial partial pressure of carbon dioxide), and an indicator for daily corticosteroid administration greater than 0.5 mg/kg body weight. Time-dependent confounders were measured irregularly for each patient. We bin them in 24-hour intervals from the time of hospitalization. In cases of multiple measurements in a 24-hour window, the clinically worst measure was used. In cases of missing baseline confounders, multiple imputation using chained equations (MICE) (van Buuren and Groothuis-Oudshoorn 2011) is used. For missing data at later time points, the last observation is carried forward. Patients are censored at their day of hospital discharge, as AKI and vital stataus were unknown after this point. A distribution of the censoring weights is given in Figure S1 in the supplementary materials.

A possible limitation of our setup is that we treat discharge as a censoring event. If patients who remain in-hospital at the end of follow-up are very different from those who are discharged, and if such differences are not explained by measured covariates, then our results may be biased. Specifically, our analyses rely on Assumption 4, which applied to this problem states that (A_t, H_t) contains all the common causes of discharge at time t and future variables measured at times $s > t$,

including future outcomes. This assumption could be violated if there is an unmeasured common cause of discharge and outcomes such as an underlying biological or clinical process not captured by the covariates listed in the above paragraph.

4.3 Results

We estimate the effect of the LMTP with the SDR estimator (on account of its enhanced robustness relative to TMLE), using a version of the `lmtp R` package (Williams et al. 2022) modified to support the competing risks setting. The nuisance functions w_t and q_t are estimated using the Super Learner ensemble modeling algorithm (van der Laan et al. 2007) via the `sl3 R` package (Coyle et al. 2022). The candidate library considered by the Super Learner included a total of 32 learning algorithms, inclusive of hyperparameter variations. The algorithms included multivariate adaptive regression splines (MARS; Friedman et al. 1991); random forests (Breiman 2001; Wright and Ziegler 2017); extreme gradient boosted trees (Friedman 2001; Chen and Guestrin 2016); lasso, ridge, and elastic net (with equal weights, i.e., $\alpha = 0.5$) regularized regression (Friedman et al. 2022); main terms generalized linear models (McCullagh and Nelder 1989; Enea 2009); generalized linear models with Bayesian priors on parameter estimates (Gelman and Hill 2006); and an unadjusted outcome mean. The estimators of w_t and q_t were computed under a Markov assumption with lag of two days to reflect the collection of laboratory results at minimum 48-hour intervals.

Note that a complete clinical understanding of the effect of IMV would require estimating its effect on death, as these interventions are meant to save lives. Figure 2 displays the SDR-based incidence difference between the two treatment conditions for both outcomes, together with 95% simultaneous confidence bands.

Inspection of Fig. 2 makes apparent the protective effect of the delay-in-intubation policy against AKI. The estimates of the incidence difference are consistently negative and decrease steadily in time; moreover, the simultaneous confidence bands do not overlap with the null value of zero except for on the first day. Hypothesis testing of the incidence difference yields adjusted p values < 0.001 for all days after the first, indicating a statistically significant protective effect of the LMTP; the Bonferroni correction was used to adjust for testing multiplicity. Based on these estimates, the delay-in-intubation policy would be expected to result in a decrease in 14-day AKI rates by approximately 3.5% (simultaneous 95% CI [2.7%, 4.3%]) relative to no intervention. Importantly, while IMV does indeed have the effect of increasing survival, its negative effect on AKI is much larger than its positive effect on survival. This raises the question of whether there are patients for whom there is no survival benefit who are being harmed by the intervention. Developing methods for identifying those patients is the subject of future methodological research.

5 Discussion

Modified treatment policies represent a recent but powerful advance in modern causal inference. These intervention regimes have a variety of applications, including (i) flexibility in the definition of effects whose identification relies on a positivity assumption that may be satisfied by design; (ii) flexibility in the definition and estimation of the effects of continuous, time-valued, and multivariate treatments; and (iii) definition of effects that depend on the natural value of treatment. Here, we have advanced the LMTP methodology, extending it to the case of survival outcomes with competing events. We have presented identification assumptions for the causal parameters, as well as a sequentially doubly robust and efficient estimation algorithm capable of leveraging the many data-adaptive regression procedures available for nuisance parameter estimation. We ensure that the estimates of the survival function remain within the parameter space (monotonic decreasing and bounded in the closed unit interval) by leveraging the monotonic projection framework of Westling and Carone (2020) and Westling et al. (2020). We have illustrated the scientific utility of our approach through application to the study of negative side effects of invasive mechanical ventilation on acute kidney injury in patients hospitalized with COVID-19, for which death by other causes serves as a competing event.

Our Proposition 2 shows that doubly robust estimation is possible for any intervention that is defined in terms of an LMTP satisfying the stated conditions. Namely, any randomized intervention defined by a known function $\mathbb{d}(a_t, h_t, \varepsilon_t)$ through either an LMTP or a general stochastic intervention accommodates doubly robust estimation, provided that: (1) either (i) A_t is a discrete random variable for all t, or (ii) A_t is a continuous random variable and the modified treatment policy $\mathbb{d}$ satisfies Condition 5, and (2) the randomizer ε_t is drawn independently across units and independently of U, and its distribution does not depend on P. Notably, this rich class contains all interventions considered by Theorem 2 of Wen et al. (2021).

A limitation of the approach taken in this paper is that we considered data collected in discrete time. While the assumption of discrete time is sensible in many applications (e.g., most clinical measures are recorded in discrete time, whether the scale is seconds, minutes, hours, or days), our methods may be of limited applicability when the time scale is truly continuous or very fine (e.g., seconds) such that many time points need to be considered. In such cases, it may be possible to coarsen the time scale into discrete components. Whether such an approach provides sensible answers and the appropriate level of coarsening needs to be assessed on a case-by-case basis based on subject-matter relevance for the application at hand.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10985-023-09606-7.

Acknowledgements NSH's contributions were supported by the National Science Foundation (Award no. DMS 2102840).

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

Andersen P, Gill RD (1982) Cox's regression model for counting processes: a large sample study. Ann Stat 10:1100–1120

Bang H, Robins JM (2005) Doubly robust estimation in missing data and causal inference models. Biometrics 61:962–973

Bavishi AA, Mylvaganam RJ, Agarwal R, Avery RJ, Cuttica MJ (2021) Timing of intubation in coronavirus disease 2019: A study of ventilator mechanics, imaging, findings, and outcomes. Crit Care Explor 3:e0415

Benkeser D, Carone M, Gilbert PB (2018) Improved estimation of the cumulative incidence of rare outcomes. Stat Med 37:280–293

Bickel PJ (1982) On adaptive estimation. Ann Stat 10:647–671

Bickel PJ, Klaassen CA, Ritov Y, Wellner JA (1997) Efficient and Adaptive Estimation for Semiparametric Models. Springer-Verlag, Berlin

Breiman L (1996) Stacked regressions. Mach Learn 24:49–64

Breiman L (2001) Random forests. Mach Learn 45:5–32

Buckley J, James I (1979) Linear regression with censored data. Biometrika 66:429–436

Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA (2010) When to start treatment? a systematic approach to the comparison of dynamic regimes using observational data. Int J Biostat. https://doi.org/10.2202/1557-4679.1212

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W (2017) Double/debiased/neyman machine learning of treatment effects. Am Econ Rev 107:261–65

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. Economet J 21:C1–C68

Cox DR (1972) Regression models and life-tables. J R Stat Soc: Ser B (Stat Methodol) 34:187–202

Coyle JR, Hejazi NS, Malenica I, Phillips RV, Sofrygin O (2022) sl3: Modern pipelines for machine learning and Super Learning. R package version 1(4):4

Díaz I, Hejazi NS (2020) Causal mediation analysis for stochastic interventions. J R Stat Soc: Ser B (Stat Methodol) 82:661–683

Díaz I, Hejazi NS, Rudolph KE, van der Laan MJ (2020) Non-parametric efficient causal mediation with intermediate confounders. Biometrika 108:627–641

Díaz I, van der Laan MJ (2011) Super learner based conditional density estimation with application to marginal structural models. Int J Biostat 7:1–20

Díaz I, van der Laan MJ (2012) Population intervention causal effects based on stochastic interventions. Biometrics 68:541–549

Díaz I, van der Laan M J (2018) Stochastic treatment regimes. In: Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies. Springer, pp. 219–232

Díaz I, Williams N, Hoffman KL, Schenck EJ (2021) Nonparametric causal effects based on longitudinal modified treatment policies. J Am Stat Assoc, 1–16

Durdevic M, Durdevic D, Riera MB, Nimkar A, Stan AC, Hasan A, Naaraayan A, Jesmajian S (2020) Progressive renal failure in patients with covid-19 after initiating mechanical ventilation: A case series. Chest 158:A2629

Enea M (2009) Fitting linear models and generalized linear models with large data sets in R. In: Statistical Methods for the Analysis of Large Datasets. pp. 411–414

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94:496–509

Friedman J, Hastie T, Tibshirani R., Narasimhan B, Tay K, Simon N, Yang J (2022) glmnet: Lasso and elastic-net regularized generalized linear models. R package version 4.1-3

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232

Friedman JH et al (1991) Multivariate adaptive regression splines. Ann Stat 19:1–67

Gelman A, Hill J (2006) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge

Gibson EA, Nunez Y, Abuawad A, Zota AR, Renzetti S, Devick KL, Gennings C, Goldsmith J, Coull BA, Kioumourtzoglou M-A (2019) An overview of methods to address distinct research questions

on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ Health 18:1–16
Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, Satlin MJ, Campion TR Jr, Nahid M, Ringel JB et al (2020) Clinical characteristics of covid-19 in new york city. N Engl J Med 382:2372–2374
Haneuse S, Rotnitzky A (2013) Estimation of the effect of interventions that modify the received treatment. Stat Med 32:5260–5277
Hasan SS, Capstick T, Ahmed R, Kow CS, Mazhar F, Merchant HA, Zaidi STR (2020) Mortality in COVID-19 patients with acute respiratory distress syndrome and corticosteroids use: a systematic review and meta-analysis. Expert Rev Respir Med 14:1149–1163
Hejazi NS, Benkeser D, Díaz I, van der Laan MJ (2022a) Efficient estimation of modified treatment policy effects based on the generalized propensity score
Hejazi NS, Benkeser DC, van der Laan MJ (2022b) haldensify: Highly adaptive lasso conditional density estimation. R package version 0.2.3
Hejazi NS, Rudolph KE, van der Laan MJ, Díaz I (2022c) Nonparametric causal mediation analysis for stochastic interventional (in) direct effects. Biostatistics in press
Hejazi NS, van der Laan MJ, Janes HE, Gilbert PB, Benkeser DC (2020) Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. Biometrics 77:1241–1253
Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. Biostatistics 1:465–480
Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biom J: J Math Methods Biosci 50:346–363
Husain-Syed F, Slutsky AS, Ronco C (2016) Lung-kidney cross-talk in the critically ill patient. Am J Respir Crit Care Med 194:402–414
Joannidis M, Forni LG, Klein SJ, Honore PM, Kashani K, Ostermann M, Prowle J, Bagshaw SM, Cantaluppi V, Darmon M et al (2020) Lung-kidney interactions in critically ill patients: consensus report of the acute disease quality initiative (ADQI) 21 Workgroup. Intensive Care Med 46:654–672
Kennedy EH (2019) Nonparametric causal effects based on incremental propensity score interventions. J Am Stat Assoc 114:645–656
Kes P, Jukić NB (2010) Acute kidney injury in the intensive care unit. Bosn J Basic Med Sci 10:S8
Klaassen CA (1987) Consistent estimation of the influence function of locally asymptotically linear estimators. Ann Stat 15:1548–1562
Luedtke AR, Sofrygin O, van der Laan MJ, Carone M (2017) Sequential double robustness in right-censored longitudinal models. arXiv preprint arXiv:1705.02459
McCullagh P, Nelder JA (1989) Generalized Linear Models. CRC Press, Boco Raton
Papoutsi E, Giannakoulis VG, Xourgia E, Routsi C, Kotanidou A, Siempos II (2021) Effect of timing of intubation on clinical outcomes of critically ill patients with COVID-19: A systematic review and meta-analysis of non-randomized cohort studies. Crit Care 25:1–9
Pearl J (2000) Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge
Perkins GD, Couper K, Connolly B, Baillie JK, Bradley JM, Dark P, De Soyza A, Gorman E, Gray A, Hamilton L et al (2020) Recovery-respiratory support: Respiratory strategies for patients with suspected or proven COVID-19 respiratory failure; continuous positive airway pressure, high-flow nasal oxygen, and standard care: A structured summary of a study protocol for a randomised controlled trial. Trials 21:1–3
Prentice RL, Kalbfleisch JD, Peterson AV Jr, Flournoy N, Farewell VT, Breslow NE (1978) The analysis of failure times in the presence of competing risks. Biometrics 34:541–554
Richardson TS, Robins JM (2013) Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper 128: 2013
Robins J, Li L, Tchetgen Tchetgen E, van der Vaart AW (2009) Quadratic semiparametric von Mises calculus. Metrika 69:227–247
Robins, JM (2000) Robust estimation in sequentially ignorable missing data and causal inference models. In Proceedings of the American Statistical Association
Robins JM, Hernán MA, Siebert U (2004) Effects of multiple interventions. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors 1:2191–2230
Robins JM, Rotnitzky A, Zhao L (1994) Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 89:846–866

Rotnitzky A, Robins J, Babino L (2017) On the multiply robust estimation of the mean of the g-functional. arXiv preprint arXiv:1705.08582
Rubin D, van der Laan MJ (2007) A doubly robust censoring unbiased transformation. Int J Biostat 3:1–21
Rudolph KE, Gimbrone C, Matthay EC, Diaz I, Davis CS, Pamplin II JR, Keyes K, Cerda M (2021) When effects cannot be estimated: redefining estimands to understand the effects of naloxone access laws. arXiv preprint arXiv:2105.02757
Rytgaard HC, van der Laan MJ (2021) One-step TMLE for targeting cause-specific absolute risks and survival curves. arXiv preprint arXiv:2107.01537
Schenck EJ, Hoffman KL, Cusick M, Kabariti J, Sholle ET, Campion TR Jr (2021) Critical care database for advanced research (cedar): An automated method to support intensive care units with electronic health record data. J Biomed Inform 118:103789
Stock JH (1989) Nonparametric policy analysis. J Am Stat Assoc 84:567–575
Thomson DA, Calligaro GL (2021) Timing of intubation in COVID-19: Not just location, location, location? Crit Care 25:1–2
Tobin MJ (2006) Principles and practice of mechanical ventilation
Tobin MJ (2020) Basing respiratory management of COVID-19 on physiological principles. Am J Respir Crit Care Med 201:1319–1320
Tobin MJ, Laghi F, Jubran A (2020) Caution about early intubation and mechanical ventilation in COVID-19. Ann Intensive Care 10:1–3
van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate imputation by chained equations in R. J Stat Softw 45:1–67
van der Laan MJ, Dudoit S, van der Vaart AW (2006) The cross-validated adaptive epsilon-net estimator. Stat Decis 24:373–395
van der Laan MJ, Petersen ML, Joffe MM (2005) History-adjusted marginal structural models and statically-optimal dynamic treatment regimens. Int J Biostat. https://doi.org/10.2202/1557-4679.1003
van der Laan MJ, Polley EC, Hubbard AE (2007) Super learner. Stat Appl Genet Mol Biol. https://doi.org/10.2202/1544-6115.1309
van der Laan MJ, Robins JM (2003) Unified Methods for Censored Longitudinal Data and Causality. Springer, New York
van der Laan MJ, Rose S (2011) Targeted Learning: Causal Inference for Observational and Experimental Data. Springer, New York
van der Laan MJ, Rose S (2018) Targeted Learning in Data Science: Causal Inference for Complex longitudinal Studies. Springer, New York
van der Laan MJ, Rubin D (2006) Targeted maximum likelihood learning. Int J Biost. https://doi.org/10.2202/1557-4679.1043
van der Vaart AW (1998) Asymptotic Statistics. Cambridge University Press, Cambridge
VanderWeele TJ, Vansteelandt S, Robins JM (2014) Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology 25:300
von Mises R (1947) On the asymptotic distribution of differentiable statistical functions. Ann Math Stat 18:309–348
Wen L, Marcus J, Young J (2021) Intervention treatment distributions that depend on the observed treatment process and model double robustness in causal survival analysis. arXiv preprint arXiv:2112.00807
Westling T, Carone M (2020) A unified study of nonparametric inference for monotone functions. Ann Stat 48:1001
Westling T, van der Laan MJ, Carone M (2020) Correcting an estimator of a multivariate monotone function with isotonic regression. Electron J Stat 14:3032–3069
Williams NT, Hoffman KL, Díaz I (2022) lmtp: Non-parametric causal effects of feasible interventions based on modified treatment policies. R package version 1:1
Wolpert DH (1992) Stacked generalization. Neural Netw 5:241–259
Wright MN, Ziegler A (2017) `ranger`: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77
Young JG, Hernán MA, Robins JM (2014) Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. Epidemiol Method 3:1–19
Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2020) A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med 39:1199–1236

Zeger SL, Liang K-Y (1992) An overview of methods for the analysis of longitudinal data. Stat Med 11:1825–1839

Zheng W, van der Laan MJ (2011) Cross-validated targeted minimum-loss-based estimation. Causal Inference for Observational and Experimental Data. Springer, In Targeted Learning, pp 459–474