Published in final edited form as:

Biometrics. 2017 September ; 73(3): 981–989. doi:10.1111/biom.12639.

Direct Estimation for Adaptive Treatment Length Policies: Methods and Application to Evaluating the Effect of Delayed PEG Insertion

Xin Lu and

Department of Biostatistics and Programming, Sanofi, Bridgewater, New Jersey 08807, U.S.A

Brent A. Johnson

Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York 14642, U.S.A

Summary

Dysphagia is a primary cause of death among patients diagnosed with amyotrophic lateral sclerosis (ALS), and percutaneous endoscopic gastrostomy (PEG) is a procedure to insert a tube into the stomach to assist or replace oral feeding. It is believed that PEG is beneficial and, generally, earlier insertion is preferable to later. However, gathering clinical evidence to support these beliefs on the use and timing of PEG is challenging because controlled clinical trials are not feasible and clinical endpoints are confounded with PEG in observational data. Moreover, the confounders are time-varying and time to PEG insertion may be only partially observed. We show how one can view this problem as an adaptive treatment length policy and propose a new estimator via g -computation. We show that our estimator is consistent and asymptotically normal for the causal estimand and explore its finite sample properties in simulation studies. Finally, using more than 10 years of data from Emory ALS clinic registry, we found no evidence to suggest that earlier PEG reduced 4-year mortality; thus, our results do not support the hypothesis and belief that initiating palliative care earlier extends life, on average. At the same, we cannot be certain that all important confounding variables are collected and observed to ensure our modeling assumptions are correct, so more work is needed to address these important end-of-life questions for ALS patients.

Keywords

Competing risks; Dynamic regime; g -computation; Informative censoring; Structural nested models

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a rare but fatal neurodegenerative disease. Its primary cause of death rises from dysphagia, which leads to undernutrition and respiratory

6. Supplementary Materials

Web Appendix A, referenced in Section 2.4, R programs implementing the proposed method and an example dataset are available with this paper at the *Biometrics* website on Wiley Online Library.

insufficiency. Like Parkinson's disease, ALS has no cure and existing treatments do not attempt to restore or repair dead or damaged nerves; rather, current treatments aim to slow the progression of the disease. The American Academy of Neurology recommends percutaneous endoscopic gastrostomy (PEG) insertion in patients diagnosed with ALS as an alternative or supplemental route for oral nutrition (Miller et al., 1999, 2009), which may improve quality of life and demonstrate survival advantages (Mazzini et al., 1995; Chio et al., 1999). At the same time, PEG insertion creates new emotional and physical burdens for the patient and the best time to initiate PEG insertion after diagnosis is still unclear (e.g. Forbes et al., 2004; Miller et al., 2009). Studying the clinical benefits of shorter versus longer time to PEG insertion in a randomized controlled trial in ALS patients is ethically infeasible. Instead, one may attempt to address the same scientific question using observational data.

Evaluating the effect of PEG on clinical outcomes using data from a clinical registry is challenging for several reasons. First, whether and when patients get PEG is not randomly assigned. Rather, providers advise patients on potential benefits and risks of PEG and then patients decide whether to move forward with palliative treatment or not. Second, some patients experience events which preclude them from inserting a PEG tube: for example, a major cardiac or stroke event which leads to severe physical debilitation or simply the end of the study. Thus, a decision to wait or move forward with PEG insertion continues from the date of diagnosis until the date of PEG insertion or until a decision-terminating event occurs, whichever comes first. The timing and presence of decision-terminating events are informative insofar as features of these events are generally correlated with the clinical endpoint. Accounting for both decision-terminating events and confounding in a mindful way are the two main statistical objectives; evaluating time to initiate palliative care that optimizes the clinical endpoint is the primary substantive objective.

We adopt a potential outcome framework (Rubin, 1974) to organize the statistical problem and achieve our goals. We define the potential outcome Y_t^* as the outcome if a randomly selected patient from the population inserted a PEG tube t months after the date of diagnosis and Y_ξ^* as the outcome if a patient experienced a decision-terminating event ξ months after the date of diagnosis. Combining these definitions, the potential outcome $Y_{t \wedge \xi}^*$ is the outcome if the wait time to initiate PEG is $t \wedge \xi$ months, i.e.,

$$Y_{t \wedge \xi}^* = Y_t^* I(\xi \geq t) + Y_\xi^* I(\xi < t), \quad (1)$$

where $a \wedge b = \min(a, b)$. In the current paper, our goal is to estimate the causal estimand $E(Y_{t \wedge \xi}^*)$ in the presence of time-dependent confounding using the observed data.

The definition of potential outcome in (1) and causal estimand $E(Y_{t \wedge \xi}^*)$ were first proposed in Johnson and Tsiatis (2004) for application to infusion trials, where infusion is started at baseline and stopped by provider at time t or by infusion-terminating event at time ξ , whichever came first, and they referred to the act of treating toward the minimum of t and ξ

as an infusion length policy or, more generally, as a *treatment length policy*. If we let $t_1 < \dots < t_K$ (K fixed and known) be an increasing sequence of target lengths that subsequently define the policies, then one can conceive of a sequence of treatment decisions whether to stop or continue treatment, provided no treatment-terminating event has taken place. In the present application for assessing when to initiate PEG, there is no treatment that begins at baseline and is later stopped, however the sequential decision making and decision-terminating events in initiating PEG parallel those in treatment length policies. After patients are diagnosed with ALS, they make decisions at $t_1, \dots, t_K$ whether to avail themselves of palliative care options, provided no decision-terminating events have occurred at the time the decision to initiate palliative care is to be made. Drawing on connections to treatment length policies (Johnson and Tsiatis, 2004, 2005) allows us to re-purpose some statistical ideas that were originally conceived for a different problem. In particular, Johnson and Tsiatis (2004; 2005) showed how treatment length policies may be viewed as a particular dynamic regime (Murphy et al., 2001) and we adopt this perspective for our application to initiating PEG in Section 2.

By now, it is well-known there exist many methods for estimating causal estimands. In this paper, we focus on comparing and contrasting two strategies for estimating $E(Y_{t \wedge \xi}^*)$: inverse probability weighting (IPW) via propensity score (e.g. Robins, 1997; Murphy et al., 2001) and sequential estimation through nested models (e.g. Robins, 1989, 1994, 1999). Both estimation approaches can account for time-varying confounding between patient characteristics and treatment assignment but do so in different ways. The former strategy models treatment assignment as a function of covariate history and then inversely weights the observed outcomes by the probability of the observed treatment length; the latter approach models potential outcomes sequentially through time as a function of covariate history on actions consistent with the dynamic regime and then integrates over observed covariate histories. Johnson and Tsiatis (2004, 2005) provide details on how to estimate $E(Y_{t \wedge \xi}^*)$ using inverse weighting and here we propose an alternative estimation procedure for the same causal estimand through nested models and Robins' (1986) g -computation method. Using g -computation for estimating causal estimands in dynamic regimes is not new (e.g., Moodie et al., 2009; Wahed, 2010; Wahed and Thall, 2013), however, the technical and computational details of g -computation for the particular decision rules and terminating events that lead to the causal estimand $E(Y_{t \wedge \xi}^*)$ has not been done. Additional comments regarding the scientific merit of this work are relegated to Section 5.

2. Methods

2.1 Time to Initiate PEG as a Dynamic Regime

To make the connection to dynamic regimes explicit, define a window $[t_0, t_K]$ where patients may initiate PEG, with $t_0 \equiv 0$ equal to the date of diagnosis and t_K the last point at which a patient could reasonably initiate PEG. Then, divide the period of treatment administration into K disjoint intervals $(t_{j-1}, t_j]$, $j = 1, \dots, K$, where $t_0 \equiv 0$. Here, the interval widths do not necessarily need to be equal. Rather than defining a single terminating event ξ on the real line as in (1), we define decision-terminating events on each interval. In particular, let C_j be the binary indicator for the presence or absence of a decision-terminating event in the

interval $(t_{j-1}, t_j], j = 1, \dots, K$; for example, $C_j = \mathbb{I}(t_{j-1} < \xi \leq t_j)$ and let $t_0, \dots, t_m$ be the decision targets of interest ($1 \leq m \leq K$). Using this notation, define the decision rule

$$d_j^{(m)}(\bar{C}_{j-1}) = \begin{cases} \text{initiate palliative care (e. g. PEG),} & \text{if } t_j = t_m \text{ and } \bar{C}_{j-1} = 0; \\ \text{no action,} & \text{otherwise,} \end{cases}$$

where $\bar{C}_j$ is the history of decision-terminating events $(C_0, \dots, C_j)$. The decision rule is consistent with treatment length policy by Johnson and Tsiatis (2004) as an individual can initiate palliative care at t_j provided no decision-terminating event has occurred prior to time t_j . If we let $\bar{d}_j^{(m)} = (d_1^{(m)}, \dots, d_j^{(m)})$ denote the history of decision rules through time t_j , then $\bar{d}_m^{(m)}$ is an example of a non-stochastic dynamic treatment regime (Murphy et al., 2001); our scientific objective is to estimate and compare mean endpoints on the K different dynamic regimes $\bar{d}_1^{(1)}, \dots, \bar{d}_K^{(K)}$.

Using notation in Murphy et al. (2001), the parameter of interest is mean outcome, $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}]$, where $E_{\bar{d}_m^{(m)}}$ is expectation under the dynamic regime $\bar{d}_m^{(m)}$ for the period t_0 to t_m and $\mathcal{G}$ are subgroups of interest. We note that $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}]$ may or may not be parametrized as a function of subgroups $\mathcal{G}$.

2.2 Likelihood

If it were possible to observe and collect potential outcomes $Y_{\bar{d}_m^{(m)}}^*$ from a random sample of individuals in the population, then $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}]$ could be estimated consistently using the sample averages stratified by $\mathcal{G}$. However, the observed endpoints are $Y \equiv Y_{\bar{A}_K}^*$, where A_j denotes binary decision to initiate palliative care or not at t_j , $j = (A_1, \dots, A_j)$ is decision history through time t_j , $j = 1, \dots, K$. Thus, $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}]$ cannot be estimated using a simple average of observed outcomes and other methods are needed.

To motivate the methods below, we first describe the observed data. We conceive of the observed data as follows, $\mathbf{Z} = \{(\mathbf{X}_0, C_0), A_1 \mathbb{I}(C_0 = 0), (\mathbf{X}_j, C_j) \mathbb{I}(\bar{C}_{j-1} = 0, A_j = 0), A_{j+1} \mathbb{I}(\bar{C}_j = 0, A_j = 0), Y, j = 1, \dots, K-1\}$, where $\mathbf{X}_0$ denotes all the covariate information at baseline; $\mathbf{X}_j$ is covariate information collected at time t_j and C_j denotes whether the potential decision-terminating event occurs in the interval $(t_j, t_{j+1}]$, $j = 0, \dots, K-1$. A visual summary of the observed data is described through the graph in Figure 1.

Patients enter the study with baseline information $\mathbf{X}_0$ and then may experience a decision-terminating event C_0 in the interval $(0, t_1]$. For patients who experience a decision-terminating event in the first interval, their observed treatment length is, without loss of generality, the right-hand endpoint t_1 . For the remaining patients who do not experience decision-terminating events, they may decide to initiate palliative care at time t_1 and this

stochastic binary decision is labeled A_1 . This decision process continues until the end of the decision window is reached, i.e. t_K .

Let $V = t_K \wedge \min\{t_j | A_j = 1, j = 1, \dots, K\}$ be the time when a participant initiates palliative care and $\tilde{\xi} = t_K \wedge \min\{t_j | C_j = 1, j = 1, \dots, K\}$ be the right-hand endpoint of the first interval where a decision-terminating event takes place; recall, $C_j = 1$ if $t_j < \xi - t_{j+1}$. These two events can then be used to define an indicator to denote the reason for terminating the decision process, $I(V < \tilde{\xi})$, the observed decision window $U = V \wedge \tilde{\xi}$, and observed covariate history, $\bar{\mathbf{X}}_U$. Then, the likelihood is

$$\text{LIK}(\mathbf{z}) = \text{LIK}_{\text{PO}}(\mathbf{z}) \text{LIK}_{\text{TA}}(\mathbf{z}), \quad (2)$$

where $\mathbf{z} = (y, u, \delta, \bar{\mathbf{x}}_u)$ and

$$\begin{aligned} \text{LIK}_{\text{PO}}(\mathbf{z}) &= \prod_{j=0}^{K-1} [h_j(\mathbf{x}_j | \bar{\mathbf{x}}_{j-1}) \{\lambda_j(\bar{\mathbf{x}}_j)\}^{c_j} \{1 - \lambda_j(\bar{\mathbf{x}}_j)\}^{1-c_j}]^{I(\bar{c}_{j-1}=0, \bar{a}_j=0)} f(y | u, \delta, \bar{\mathbf{x}}_u), \\ \text{LIK}_{\text{TA}}(\mathbf{z}) &= \prod_{j=1}^K [\{\pi_j(\bar{\mathbf{x}}_{j-1})\}^{a_j} \{1 - \pi_j(\bar{\mathbf{x}}_{j-1})\}^{1-a_j}]^{I(\bar{c}_{j-1}=0, \bar{a}_{j-1}=0)}, \end{aligned}$$

$h_j(\mathbf{x} | \bar{\mathbf{x}}_{j-1}) = P(\mathbf{X}_j \in \mathbf{x} | \bar{\mathbf{X}}_{j-1} = \bar{\mathbf{x}}_{j-1}, \bar{\mathbf{C}}_{j-1} = 0, j = 0)$, $\lambda_j(\bar{\mathbf{x}}_j) = P(C_j = 1 | \bar{\mathbf{X}}_j = \bar{\mathbf{x}}_j, \bar{\mathbf{C}}_{j-1} = 0, j = 0)$, $\pi_j(\bar{\mathbf{x}}_{j-1}) = P(A_j = 1 | \bar{\mathbf{X}}_{j-1} = \bar{\mathbf{x}}_{j-1}, \bar{\mathbf{C}}_{j-1} = 0, j = 0)$, $f(y | u, \delta, \bar{\mathbf{x}}_u) = \lim_{\epsilon \downarrow 0} \epsilon^{-1} P(y < Y - y + \epsilon | U = u, \delta, \bar{\mathbf{X}}_U = \bar{\mathbf{x}}_u)$ and $h_0 = h_0(\mathbf{x}_0 | \mathcal{G})$. The likelihood is split between LIK_{PO} and LIK_{TA} , with the former representing probability models for intermediate outcomes, covariates, and primary outcome and the latter a function of treatment assignment. The treatment assignment probabilities $\pi_j(\bar{\mathbf{x}}_{j-1})$ can be used to define a propensity score which can subsequently be used to propose an estimator of $E_{\bar{d}_m^{(m)}}[Y | \mathcal{G}]$ via inverse weighting. The details of the inverse weighting approach are given in Johnson and Tsiatis (2004) and not reviewed here. In the sequel, we estimate the causal estimand $E_{\bar{d}_m^{(m)}}[Y | \mathcal{G}]$ without modeling treatment assignment but rather through parameters contained in LIK_{PO} , i.e. f_Y , λ_0 , h_0 , ..., λ_{K-1} , h_{K-1} .

2.3 The causal estimand

Let a_j represent a patient's decision to initiate palliative care at time t_j , provided the decision window is still open at time t_j , and $\mathbf{a}_j = (a_1, \dots, a_j)$ be the history of treatment decisions through time t_j . Now, let $\mathcal{A}_m$ be the complete collection of all possible histories of decisions through time t_m . If we set $\mathbf{L}_j = (C_j, \mathbf{X}_j)$ and $\bar{\mathbf{L}}_j = (\mathbf{L}_0, \dots, \mathbf{L}_j)$, then following Murphy et al. (2001, p. 1413, eqn. 4.4), we have

$$E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}=g] = \sum_{\bar{\mathbf{a}}_m \in \mathcal{A}_m} \int_{\bar{\mathbf{I}}_{m-1}} \int_y y f_Y(y|\bar{\mathbf{a}}_m, \bar{\mathbf{I}}_{m-1}) \prod_{j=1}^m I\{a_j = d_j^{(m)}(\bar{\mathbf{c}}_{j-1})\} \prod_{j=1}^{m-1} f_j(\mathbf{l}_j|\bar{\mathbf{a}}_j, \bar{\mathbf{I}}_{j-1}) f_0(\mathbf{l}_0|g) dy d\mu(\bar{\mathbf{I}}_{m-1}), \quad (3)$$

where $f_j(\mathbf{l}_j|\bar{\mathbf{I}}_{j-1}) = [h_j(\mathbf{x}_j|\bar{\mathbf{x}}_{j-1})\{\lambda_j(\bar{\mathbf{x}}_j)\}^{c_j}\{1-\lambda_j(\bar{\mathbf{x}}_j)\}^{1-c_j}]^{I\{\bar{\mathbf{c}}_{j-1}=0, j \neq 0\}}$, $f_0(\mathbf{l}_0|g) = \lambda_0(\mathbf{x}_0|g)h_0(\mathbf{x}_0|g)$, and $\mu(\bar{\mathbf{I}}_{m-1})$ is the dominating measure for $\bar{\mathbf{I}}_{m-1}$. The expression in (3) is Robin's (1986) g -computation formula for dynamic regimes applied to treatment length policies. Note, that unlike (4.4) in Murphy et al., the products on the right-hand side of (3) go to time t_m rather than t_K because, as mentioned earlier, patients continuously treated beyond time t_m cannot possibly be following the regime $\bar{d}_m^{(m)}$ and so $I\{a_j = d_j^{(m)}(\bar{\mathbf{c}}_{j-1})\} = 0$ for $t_j > t_m$.

Expression (3) is general and allows for broad definitions of decision rules; at the same time, expression (3) obfuscates the competing risks in treatment decisions issue raised earlier. For example, the outcome Y is a function of data only through the observed treatment length U , which may be less than t_m , then the notation $f_Y(y|\bar{\mathbf{a}}_m, \bar{\mathbf{I}}_{m-1})$ masks this fact. We offer an alternative formula in (4) which we believe is more intuitive. Recall, for regime $\bar{d}_m^{(m)}$, either a decision-terminating event occurred in one of m intervals prior to t_m or the provider stopped treatment at t_m . By considering each of the m possible options, we provide the following expression for the causal estimand:

$$E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}=g] = \int_{\bar{\mathbf{x}}_{m-1}} b_1(t_m, \bar{\mathbf{x}}_{m-1}) \prod_{j=1}^{m-1} [\{1-\lambda_j(\bar{\mathbf{x}}_j)\}h_j(\mathbf{x}_j|\bar{\mathbf{x}}_{j-1})] [\{1-\lambda_0(\mathbf{x}_0|g)\}h_0(\mathbf{x}_0|g)] d\bar{\mathbf{x}}_{m-1} \\ + \sum_{j=1}^{m-1} \left\{ \int_{\bar{\mathbf{x}}_j} b_0(t_{j+1}, \bar{\mathbf{x}}_j) \lambda_j(\bar{\mathbf{x}}_j) \prod_{l=1}^j [\{1-\lambda_l(\bar{\mathbf{x}}_l)\}h_l(\mathbf{x}_l|\bar{\mathbf{x}}_{l-1})] [\{1-\lambda_0(\mathbf{x}_0|g)\}h_0(\mathbf{x}_0|g)] d\bar{\mathbf{x}}_j \right\} \\ + \int_{\mathbf{x}_0} b_0(t_1, \mathbf{x}_0) \lambda_0(\mathbf{x}_0|g) h_0(\mathbf{x}_0|g) d\mathbf{x}_0 \quad (4)$$

where $b_g(u, \bar{\mathbf{x}}_u) = \int y f_Y(y|u, \delta, \bar{\mathbf{x}}_u) dy$ using our definition of f_Y in Section 2.2. In contrast to expression (3), expression (4) makes explicit the fact that Y depends on information collected up to the observed treatment length U but not afterwards.

Finally, we remark that the validity of the equalities in (3) and (4) rely on two assumptions: the stable unit treatment value assumption (SUTVA; Rubin, 1978) and the sequential randomization assumption (Robins, 1986). Sequential randomization assumption asserts that, conditional on covariate history, decisions to initiate PEG are not influenced by future potential outcomes while SUTVA assumes that patients' outcomes are not influenced by the outcomes of others. However, as highlighted by Murphy et al. (2001), the right-hand side of (3) or (4) can be useful statistics irrespective of these assumptions.

2.4 A direct estimator

The proposed direct estimators approximate each of the conditional probability density and hazard functions h_j , λ_j , f_Y , $j = 0, \dots, m-1$, via statistical models and then evaluate the multi-dimensional integral in (4); i.e.,

$$\begin{aligned}
 &= \int_{\bar{\mathbf{x}}_{m-1}} \hat{b}_1(t_m, \bar{\mathbf{x}}_{m-1}) \prod_{j=1}^{m-1} \left[\{1 - \hat{\lambda}_j(\bar{\mathbf{x}}_j)\} \hat{h}_j(\mathbf{x}_j | \bar{\mathbf{x}}_{j-1}) \right] \left[\{1 - \hat{\lambda}_0(\mathbf{x}_0 | g)\} \hat{h}_0(\mathbf{x}_0 | g) \right] d\bar{\mathbf{x}}_{m-1} \\
 &+ \sum_{j=1}^{m-1} \left\{ \int_{\bar{\mathbf{x}}_j} \hat{b}_0(t_{j+1}, \bar{\mathbf{x}}_j) \hat{\lambda}_j(\bar{\mathbf{x}}_j) \prod_{l=1}^j \left[\{1 - \hat{\lambda}_l(\bar{\mathbf{x}}_l)\} \hat{h}_l(\mathbf{x}_l | \bar{\mathbf{x}}_{l-1}) \right] \left[\{1 - \hat{\lambda}_0(\mathbf{x}_0 | g)\} \hat{h}_0(\mathbf{x}_0 | g) \right] d\bar{\mathbf{x}}_j \right\} \\
 &\quad + \int_{\mathbf{x}_0} \hat{b}_0(t_1, \mathbf{x}_0) \hat{\lambda}_0(\mathbf{x}_0 | g) \hat{h}_0(\mathbf{x}_0 | g) d\mathbf{x}_0
 \end{aligned}
 \tag{5}$$

where $\hat{b}_0$, $\hat{b}_1$, $\hat{\lambda}_j$ and $\hat{h}_j$ are consistent estimators of b_0 , b_1 , λ_j and h_j respectively. For example, if $\hat{b}_0$, $\hat{b}_1$, $\hat{\lambda}_j$ and $\hat{h}_j$ were estimated nonparametrically, then substituting these estimators in (5) would yield a completely nonparametric estimator of $E_{\bar{d}_m^{(m)}}[Y | \mathcal{G} = g]$. In the sequel, we provide details for a semi-parametric approach.

First, we assume that the time-dependent covariates follow a first-order Markov model, $h_j(\mathbf{x}_j | \bar{\mathbf{x}}_{j-1}; \boldsymbol{\theta}_j) \equiv h_j(\mathbf{x}_j | \mathbf{x}_{j-1}; \boldsymbol{\theta}_j)$, $j = 0, \dots, K-1$ and $\boldsymbol{\theta}_j$ is an unknown parameter vector and $\boldsymbol{\theta}_0 = \text{vec}\{E(\mathbf{X}_0), \text{cov}(\mathbf{X}_0)\}$, with $\text{vec}(\boldsymbol{\Omega})$ denoting a vectorized list of the unique elements in $\boldsymbol{\Omega}$. Next, we assume that the probability of a decision-terminating event follows a discrete hazard model where the hazard depends on the covariate history through the most recent covariate value only, i.e.

$$\lambda_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta}) = F(\eta_j + \mathbf{X}_0^T \boldsymbol{\eta}_{X_0} + \mathbf{X}_j^T \boldsymbol{\eta}_{X_t}) \tag{6}$$

and $R(p) = \{1 + \exp(-p)\}^{-1}$ and $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{K-1}, \boldsymbol{\eta}_{X_0}^T, \boldsymbol{\eta}_{X_t}^T)^T$. Finally, we assume the outcome Y given $(U, \bar{\mathbf{X}}_U)$ belongs to the exponential family of distributions. In the simulations, we assume Y follows a Gaussian linear model with mean,

$$E(Y | U, \Delta, \bar{\mathbf{X}}_U; \boldsymbol{\alpha}) = \alpha_0 + \log(U) \alpha_U + \Delta \alpha_\Delta + \mathbf{X}_0^T \boldsymbol{\alpha}_{X_0} + \mathbf{X}_{[U]}^T \boldsymbol{\alpha}_{X_t}, \tag{7}$$

and variance σ_Y^2 , where $\Gamma U = \max\{j | t_j \leq U\}$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_U, \alpha_\Delta, \boldsymbol{\alpha}_{X_0}^T, \boldsymbol{\alpha}_{X_t}^T)^T$. These models can be fit using routine statistical software and the estimated regression parameters returned as standard output.

Using these statistical models, $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{K-1})^T$, $\boldsymbol{\eta}$ and $\boldsymbol{a}$ completely parameterize the partial likelihood LIK_{PO} in (2) and define h_j , λ_j , b_1 and b_0 in (5). Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_0, \dots, \hat{\boldsymbol{\theta}}_{K-1})^T$, $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{a}}$ be consistent estimators of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{a}$, respectively. Then, substituting the (semi-) parametric predictions $\hat{h}_j$, $\hat{\lambda}_j$ and $\hat{b}_1$ and $\hat{b}_0$ into (5) and evaluating the integral leads to an estimator of $E_{\tilde{d}_m^{(m)}}[Y|\mathcal{G}]$. Under suitable regularity conditions, including that every distribution and model is correctly specified, performing this sequential estimation procedure will result in a consistent estimator of $E_{\tilde{d}_m^{(m)}}[Y|\mathcal{G}]$. (More details regarding deriving direct estimators from causal estimand could be found in Web Appendix A in the Supplementary Materials.) We investigate two estimation strategies below.

2.4.1 Monte Carlo integration—After the parameters are estimated, they can be used to construct an estimate of $E[Y_{\tilde{d}_m^{(m)}}|\mathcal{G}]$ via Monte Carlo integration. Let r denote the r th random sample, $r = 1, \dots, R$, where R is a large number. Fix t_m , the target time to initiate PEG. Draw $\mathbf{X}_{0r} \sim h_0(\mathbf{x}_0|\mathcal{G}; \hat{\boldsymbol{\theta}}_0)$ and $C_{0r} \sim \text{Bern}\{\lambda_0(\mathbf{X}_{0r}; \hat{\boldsymbol{\eta}})\}$. If $C_{0r} = 1$, stop; if $C_{0r} = 0$, continue. Draw $\mathbf{X}_{1r} \sim h_1(\mathbf{x}_1|\mathbf{X}_{0r}; \hat{\boldsymbol{\theta}}_1)$ and $C_{1r} \sim \text{Bern}\{\lambda_1(\mathbf{X}_{1r}; \hat{\boldsymbol{\eta}})\}$. If $C_{1r} = 1$, stop; if $C_{1r} = 0$, continue. Continue to simulate $(\mathbf{X}_{jr}, C_{jr})$ until the interval just prior to the target time t_m . If no decision-terminating event happens in this interval, i.e. $C_{(m-1)r} = 0$, then the patient initiates PEG at t_m , in which case $U_r = t_m$ and $\tilde{d}_r = 1$. In all other cases where a decision-terminating event preceded t_m , $U_r = \tilde{t}_r$ and $\tilde{d}_r = 0$. Then, the Monte Carlo estimator is

$$\hat{\mu}_{\text{MC}}^{(m)} = R^{-1} \sum_{r=1}^R E(Y|U_r, \Delta_r, \mathbf{X}_{0r}, \mathbf{X}_{[U_r]}; \hat{\boldsymbol{\theta}}), \text{ using the model in (7) for } E(Y|U_r, \Delta_r, \mathbf{X}_{0r}, \mathbf{X}_{[U_r]}; \boldsymbol{a}).$$

2.4.2 Direct Prediction—A second estimator attempts to relax some of the distributional assumptions in $\hat{\mu}_{\text{MC}}^{(m)}$ and speed up the computations. Rather than modeling the covariate history via $h_j(\mathbf{x}_j|\mathbf{x}_{j-1})$, $j = 0, \dots, m-1$, one can rewrite (4) as a function of the observed covariate histories $\bar{\mathbf{X}}_j|(\mathbf{U} \leq t_j)$ without having to model h_j . In Supplementary Material, we show how (4) may be written as

$$\lim_{n \rightarrow \infty} \mathbb{P}_n \left[b_1(t_m, \bar{\mathbf{X}}_{m-1}) Q_{m-1}(\bar{\mathbf{X}}_{m-1}) I(U > t_{m-1}) + \sum_{j=0}^{m-1} b_0(t_{j+1}, \bar{\mathbf{X}}_j) q_j(\bar{\mathbf{X}}_j) I(U > t_j) \right], \quad (8)$$

where $Q_{m-1}(\bar{\mathbf{X}}_{m-1}) = \prod_{j=0}^{m-1} \{1 - \lambda_j(\bar{\mathbf{X}}_j)\}$, $q_j(\bar{\mathbf{X}}_j) = \lambda_j(\bar{\mathbf{X}}_j) \prod_{l=0}^{j-1} \{1 - \lambda_l(\bar{\mathbf{X}}_l)\}$. Using our model in (6) for the cause-specific hazard function $\lambda_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta})$ and (7) for the outcome models $b_1(t_m, \bar{\mathbf{X}}_{m-1}; \boldsymbol{a})$ and $b_0(t_j, \bar{\mathbf{X}}_j; \boldsymbol{a})$ and estimated parameters $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{a}}$, we can form predicted values for the remaining unknown parameters in (8). If we replace $\lambda_j(\bar{\mathbf{X}}_j)$ with $\lambda_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta})$ in the definitions of $Q_{m-1}(\bar{\mathbf{X}}_{m-1}; \boldsymbol{\eta})$ and $q_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta})$ in (8) and subsequently replace $\boldsymbol{\eta}$ and $\boldsymbol{a}$ with $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{a}}$, respectively, then our *direct prediction* estimator is given by the expression

$$\hat{\mu}_{\text{DP}}^{(m)} = \mathbb{P}_n \left[b_1(t_m, \bar{\mathbf{X}}_{m-1}; \hat{\boldsymbol{\alpha}}) Q_{m-1}(\bar{\mathbf{X}}_{m-1}; \hat{\boldsymbol{\eta}}) I(U > t_{m-1}) + \sum_{j=0}^{m-1} b_0(t_{j+1}, \bar{\mathbf{X}}_j; \hat{\boldsymbol{\alpha}}) q_j(\bar{\mathbf{X}}_j; \hat{\boldsymbol{\eta}}) I(U > t_j) \right].$$

Under the stronger assumption that there is confounding through baseline covariates $\mathbf{X}_0$ only, the estimator $\hat{\mu}_{\text{DP}}^{(m)}$ reduces to a discrete form of the estimator proposed recently by Lu and Johnson (2015).

3. The Effect of Delayed PEG Insertion in ALS patients

Now, we apply the methods in Section 2 to compare 4-year mortality rates across a range of PEG insertion times using the same data as Traxinger et al. (2013) from the Emory ALS Clinic. The dataset includes baseline demographic variables (i.e. gender, age at diagnosis), and clinical variables (i.e. bulbar or spinal site of onset, time from diagnosis date) as well as time-dependent forced vital capacity (FVC) on $n = 710$ deceased patients from the Emory ALS Clinic Registry between 1997 and 2011. In this cohort, 56% were male, site of onset was 69% bulbar, the average age at diagnosis was 61 years, the overall 4-year mortality rate was 90%. 406 patients died without ever having received PEG; if death occurred prior to 4 years, the date of death was also the date of the decision-terminating event. Using the Kaplan-Meier estimator, the median time to initiate PEG from the date of diagnosis was 6.5 months and median time to death was 16 months from diagnosis date. Additional details may be found elsewhere (Traxinger et al., 2013).

For our method, we coarsened the decision window into intervals. The intervals $(t_j, t_{j+1}]$ were defined by the interval endpoints 0, 3, 9, 18, and 36 months and chosen for convenience. We considered increasing the number of intervals but the sparsity of decision-terminating events over the decision window limited the complexity of the statistical model. 32% of the patients were missing time-dependent FVC. To overcome this deficiency, we performed our data analysis over five multiply imputed data sets using the mice package in R. In particular, the coefficient and standard error estimates (generated from 1000 bootstrap iterations) are pooled over the five data sets using Rubin's (2004) rules. Standard error estimates for the IPW estimator on a single imputed data set were computed using the estimator for the asymptotic covariance given in Johnson and Tsiatis (2004); the standard errors over five multiply imputed data sets were combined through Rubin's rules.

The parameter estimates from the intermediate models in our direct estimators are given in Table 1. We find that patients seems to have relatively lower hazard of experiencing decision-terminating events from 3 to 9 months compared to the first 3 months. The results also suggest that older patients with spinal site of onset and lower FVC have elevated risk of decision-terminating events. In the conditional outcome model, age and FVC are associated with 4-year mortality. Interpreting the association between 4-year mortality and PEG insertion directly from the results in Table 1 is complex because of confounding by indication. Fewer patients are at-risk of initiating PEG because they have already died; alternatively, the only patients eligible to initiate PEG later in the study are those patients who are still alive. This motivates our methods which leads to the results in Table 2.

Based on the results in Table 2, we conclude that delaying PEG insertion leads to lower 4-year mortality. Despite the fact that the direct and IPW estimators are based on different models and assumptions, their point estimates in Table 2 are rather similar. The IPW estimator using baseline covariates only yields similar estimates to estimates that use both baseline and time-dependent covariate information, indicating that FVC may not add a tremendous amount of new information beyond the baseline adjustment. Overall, our direct estimators are generally more precise than IPW: standard error estimates from direct estimators are roughly 17% smaller than IPW for t_j greater than 3 months and similar to IPW estimators at $t_m = 3$. The agreement between inverse weighting and direct estimation provides some heuristic validity to our overall conclusions. On the other hand, the naïve crude proportions deviates from the other estimates in later time points, suggesting they do not accurately account for decision-terminating events, confounding, or both.

4. Simulation studies

Simulation studies are conducted to assess the small sample performance of the proposed methods and to contrast with IPW estimators by Johnson and Tsiatis (2004). We begin by simulating two independent baseline covariates, X_{01} and X_{02} , as standard normal random variables and independently generating one time-dependent covariate $W_0 \sim \mathcal{N}(4, 1)$. Set $X_t \equiv W_0$. Then, simulate the potential decision-terminating event C_0 in $(0, t_1]$ as $\text{Bern}\{\lambda_0(\bar{\mathbf{X}}_0; \boldsymbol{\eta})\}$, where $\bar{\mathbf{X}}_0 = (\mathbf{X}_0^T, W_0)^T$, $\mathbf{X}_0 = (X_{01}, X_{02})^T$ and $\lambda_0(\bar{\mathbf{X}}_0; \boldsymbol{\eta})$ is given in (6). If $C_0 = 1$, then terminate the decision process, set $(U, \Delta) = (\tilde{\xi}, 0)$ and skip to model Y in (9); otherwise, continue. If $C_0 = 0$, then simulate A_1 as Bernoulli random variable with success probability $\pi_1(\bar{\mathbf{X}}_0; \boldsymbol{\gamma})$. If $A_1 = 1$, initiate PEG at t_1 , set $(U, \Delta) = (t_1, 1)$ and skip to model Y in (9); otherwise, continue. If $A_1 = 0$, then simulate the time-dependent covariate $X_t \equiv W_1 \sim \mathcal{N}(W_0, 1)$ and the potential decision-terminating event C_1 in $(t_1, t_2]$ as $\text{Bern}\{\lambda_1(\bar{\mathbf{X}}_1; \boldsymbol{\eta})\}$, with $\bar{\mathbf{X}}_1 = (\mathbf{X}_0^T, W_1)^T$. If $C_1 = 1$, stop the decision process, set $(U, \Delta) = (\tilde{\xi}, 0)$ and skip to model Y in (9); otherwise, continue. This process continues sequentially for all K stages. Following Johnson and Tsiatis (2004), if there is no decision-terminating event in the $(K - 1)$ st stage, then PEG is initiated at t_K with probability one.

For our simulation studies, we assume the outcome model is linear in observed variables, i.e.

$$Y = 1.5 + \log(U) - 2\Delta + X_{01} + 0.5X_{02} + 0.5X_t + \sigma\varepsilon, \quad (9)$$

and ε is an independent standard normal random variable. The treatment assignment model at time t_j is governed by the cause-specific hazard, $\pi_j(\bar{\mathbf{X}}_{j-1}; \boldsymbol{\gamma}) = F(\gamma_{I_j} + \mathbf{X}_0^T \boldsymbol{\gamma}_{X_0} + \mathbf{X}_t^T \boldsymbol{\gamma}_{X_t})$, with parameters $\boldsymbol{\gamma}_I = (\gamma_{I1}, \gamma_{I2}, \gamma_{I3})$, $\boldsymbol{\gamma}_{X_0} = (-0.2, -0.2)^T$, $\boldsymbol{\gamma}_{X_t} = -0.8$, $\mathbf{X}_t$ is the scalar X_t , and $F(p)$ is defined in (6). Similarly, the decision-terminating events follow a discrete hazards model with cause-specific hazard $\lambda_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta})$, $\boldsymbol{\eta}_I = (\eta_{I0}, \eta_{I1}, \eta_{I2}, \eta_{I3})^T$, $\boldsymbol{\eta}_{X_0} = (-0.1, -0.1)^T$, and $\boldsymbol{\eta}_{X_t} = -0.9$. The data simulation procedure is repeated independently for n subjects.

Now, to control the proportion of decision-terminating events, we consider two different vectors of intercept parameters, $\boldsymbol{\eta}_I$ and $\boldsymbol{\gamma}_I$ for the cause-specific hazard functions. In Scenario 1, we set $\boldsymbol{\gamma}_I = (0.5, 1, 1.5)^T$ and $\boldsymbol{\eta}_I = (0.65, 1.3, 1.95, 2.6)^T$, and leads to a decision-terminating events proportion of approximately 40% across all intervals. For Scenario 2, we set $\boldsymbol{\gamma}_I = (0.1, 0.2, 0.3)^T$ and $\boldsymbol{\eta}_I = (1, 2, 3, 4)^T$, and leads to an event proportion of roughly 65% overall. For each simulation scenario, we considered sample sizes of $n = 300$ and $n = 600$ subjects.

The direct estimators, $\hat{\mu}_{DP}^{(m)}$ and $\hat{\mu}_{MC}^{(m)}$, and IPW estimators were computed for each Monte Carlo data sets. The estimator $\hat{\mu}_{MC}^{(m)}$ is based on $R = 100,000$ samples. The IPW estimators are computed under the assumption that treatment assignment depends on time-dependent covariates and that treatment assignment depends on baseline covariates only. Standard error estimates are computed using bootstrap resampling. We experiment with 100 and 500 bootstrap resamples and find that standard error estimates are nearly identical. The true population parameter $E_{\bar{d}_m^{(m)}}[Y]$ is a complicated integral but may be computed via numerical integration. To operationalize Monte Carlo integration, for example, we repeat the data generating steps exactly as described above except we force patients to initiate PEG at time t_m if no decision-terminating event occurred prior to t_m (see, Johnson and Tsiatis, 2004). We summarize the results of our simulation studies through four statistics: Monte Carlo bias, the simulation standard error (SSE), standard error estimate (SEE), and the empirical coverage probability (ECP) is the proportion of Monte Carlo data sets where a Wald-type 95% confidence interval contains the true value. Our simulation results from 1000 Monte Carlo datasets are displayed in Table 3.

The results show that both approaches in direct estimation work well with bias close to zero. The IPW estimator that models treatment assignment using time-dependent covariates is unbiased while the estimator that ignores time-dependent confounding is biased. The bootstrap resampling scheme for the direct estimators produces standard error estimates that match very well the true sampling error. The sandwich estimator for the IPW estimator appears to match reasonably well when propensity score depends on baseline confounders only but is too small in the presence of time-dependent confounders. Consequently, the direct estimators have empirical coverage probability close to the nominal level while the IPW estimators generally have low coverage probability.

As the proportion of decision-terminating events increases from 40% to 65%, the IPW estimators exhibit larger finite sample bias and variance whereas the bias and variance remained about the same for the direct estimators. In general, when comparing direct and IPW estimators, our results suggest a 3- to 4-fold increase in the standard error of the parameter estimates when all modeling assumptions for both estimators are true. Comparing computational strategies for the direct estimators, the savings in computational cost of the direct prediction method compared to Monte Carlo integration is substantial and is highly recommended.

5. Discussion

In this article, we proposed statistical methods for estimating the clinical effect of delayed PEG insertion, a common palliative treatment option for patients diagnosed with amyotrophic lateral sclerosis. Objectively summarizing the clinical benefits or risks of PEG insertion is challenging because patients that opt for PEG at a later calendar time probably have slower disease progression and also live longer. Evaluating the effect of PEG insertion on clinical endpoints is subject to confounding by indication. The framework we adopted views time to PEG insertion as a type of treatment length policy (Johnson and Tsiatis, 2004, 2005) and then we used techniques from the dynamic regime literature (Murphy et al., 2001) to solve the estimation problem. In theory, the proposed methods allow for unbiased comparisons of mean outcome between PEG insertion at different target time points after date of diagnosis, even in the presence of informative dropout, provided that a rich set of time-dependent confounders makes the sequential randomization assumption plausible.

We proposed a new direct estimator for the mean potential outcome for subjects following a *treatment length policy*, as defined by Johnson and Tsiatis (2004, 2005). Our method adjusts for time-varying confounders and potential treatment-terminating events using a principle of g -computation and sequential modeling of the potential outcome data. Although g -computation is not itself a novel concept, our estimator for the treatment length problem differs fundamentally from earlier work by Johnson and Tsiatis (2004; 2005) because it does not model treatment assignment. Readers may find interesting the role that decision-terminating events play in our framework and the manner in which they compete with patient decisions and how this affects the sequential estimation procedure.

Finally, the direct estimators proposed here extend our methods described in Lu and Johnson (2015) from baseline confounding only to the setting where time-dependent confounding may exist. To facilitate this development, we coarsened the time axis and assumed that patient decisions take place at K distinct time points $t_1, \dots, t_K$ (see also, Johnson and Tsiatis, 2004). In our companion work, Lu and Johnson (2015) did not coarsen time and made no simplifying assumption about the treatment assignment mechanism and, from that perspective, their estimator is somewhat more general than what is proposed here. But in the presence of time-dependent confounding, the estimator by Lu and Johnson (2015) is not applicable. Developing an estimation procedure that can estimate $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}]$ in the presence of time-dependent confounding and remains unchanged regardless of whether treatment assignment is continuous or discrete is the subject of ongoing research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the comments of the associate editor and two anonymous referees that improved our writing and communication of ideas. Dr. Johnson was supported in part by the University of Rochester Center for AIDS Research grant P30AI078498 (NIH/NIAID) and the University of Rochester School of Medicine and Dentistry Clinical & Translational Science Institute. The authors thank the participants in the Emory University ALS Clinic registry as well as Dr. Jonathan Glass and Ms. Meraida Polak for providing the data.

References

- Chio A, Finocchiaro E, Meineri P, Bottacchi E, Schiffer D, et al. Safety and factors related to survival after percutaneous endoscopic gastrostomy in als. *Neurology*. 1999; 53:1123–1123. [PubMed: 10496278]
- Forbes RB, Colville S, Swingler RJ, Group SMNDR, et al. Frequency, timing and outcome of gastrostomy tubes for amyotrophic lateral sclerosis/motor neurone disease. *Journal of neurology*. 2004; 251:813–817. [PubMed: 15258782]
- Johnson BA, Tsiatis AA. Estimating mean response as a function of treatment duration in an observational study, where duration may be informatively censored. *Biometrics*. 2004; 60:315–23. [PubMed: 15180656]
- Johnson BA, Tsiatis AA. Semiparametric inference in observational duration-response studies, with duration possibly right-censored. *Biometrika*. 2005; 92:605–618.
- Lu X, Johnson BA. Direct estimation of the mean outcome on treatment when treatment assignment and discontinuation compete. *Biometrika*. 2015; 102:797–807.
- Mazzini L, Corra T, Zaccala M, Mora G, Del Piano M, Galante M. Percutaneous endoscopic gastrostomy and enteral nutrition in amyotrophic lateral sclerosis. *Journal of neurology*. 1995; 242:695–698. [PubMed: 8568533]
- Miller RG, Jackson CE, Kasarskis EJ, England J, Forsheew D, Johnston W, Kalra S, Katz J, Mitsumoto H, Rosenfeld J, et al. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology. *Neurology*. 2009; 73:1218–1226. [PubMed: 19822872]
- Miller RG, Rosenberg JA, Gelinas D, Mitsumoto H, Newman D, Sufit R, Borasio GD, Bradley WG, Bromberg MB, Brooks BR, et al. Practice parameter: The care of the patient with amyotrophic lateral sclerosis (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology. *Neurology*. 1999; 52:1311–1311. [PubMed: 10227612]
- Moodie EEM, Platt RW, Kramer MS. Estimating response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association*. 2009; 104:155–165.
- Murphy SA, van der Laan MJ, Robins JM. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96:1410–23. [PubMed: 20019887]
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods — application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
- Robins JM. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*. 1989; 113:159.
- Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods*. 1994; 23:2379–2412.
- Robins JM. Marginal structural models. 1997 Proceedings of the American Statistical Association, Section on Mayesian Statistical Science. 1997:1–10.
- Robins JM. Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. *Computation, causation, and discovery*. 1999:349–405.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.
- Rubin DB. Bayesian inference for causal effects. *The Annals of Statistics*. 1978; 6:34–58.
- Rubin, DB. Multiple imputation for nonresponse in surveys. Vol. 81. John Wiley & Sons; 2004.
- Traxinger K, Kelly C, Johnson BA, Lyles RH, Glass JD. Prognosis and epidemiology of amyotrophic lateral sclerosis analysis of a clinic population, 1997–2011. *Neurology: Clinical Practice*. 2013; 3:313–320. [PubMed: 24195020]
- Wahed AS. Inference for two-stage adaptive treatment strategies using mixture distributions. *Applied Statistics*. 2010; 59:1–18.

Wahed AS, Thall PF. Evaluating joint effects of induction-salvage treatment regimes on overall survival in acute leukemia. *Applied Statistics*. 2013; 62:67–83. [PubMed: 24014891]

Appendix

Operating characteristics

In this section, we show that the estimator $\hat{\mu}_{\text{DP}}^{(m)}$ in (8) is a partial M -estimator; that is, $\hat{\mu}_{\text{DP}}^{(m)}$ is one element of a longer vector of estimated parameters that can be written as the solution to a system of estimating equations. Let $E_{\bar{d}_m^{(m)}}[Y|\mathcal{G}=g]=\mu_m$ and define the estimating function for μ_m i.e.

$$\psi_{\mu}(\mathbf{Z}) = \mu_m - \{b_1(t_m, \bar{\mathbf{X}}_{m-1}; \boldsymbol{\alpha})Q_{m-1}(\bar{\mathbf{X}}_{m-1}; \boldsymbol{\eta})I(U > t_{m-1}) + \sum_{j=0}^{m-1} b_0(t_{j+1}, \bar{\mathbf{X}}_j; \boldsymbol{\alpha})q_j(\bar{\mathbf{X}}_j; \boldsymbol{\eta})I(U > t_j)\}$$

. Next, define the estimating functions for the parameters in the discrete hazards model, say,

$$\psi_{\boldsymbol{\eta}}(\mathbf{Z}) = (\psi_{\eta_0}, \dots, \psi_{\eta_{K-1}}, \psi_{\eta_X}^T)^T, \psi_{\eta_l} = I(U = t_{l+1}, \Delta = 0) - F(\eta_l + \bar{\mathbf{X}}_l^T \boldsymbol{\eta}_X)I(U \geq t_{l+1}), l = 0,$$

$$\dots, K-1, \psi_{\eta_X} = \sum_{j=1}^{j=K} \{I(U = t_j, \Delta = 0) - F(\eta_{j-1} + \bar{\mathbf{X}}_{j-1}^T \boldsymbol{\eta}_X)I(U \geq t_j)\} \bar{\mathbf{X}}_{j-1} \text{ and define}$$

$$\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{Z}) = (Y - \boldsymbol{\Xi}^T \boldsymbol{\alpha}) \boldsymbol{\Xi}, \boldsymbol{\Xi} = [1, \log(U), \Delta, \bar{\mathbf{X}}_U^T]^T. \text{ Now, let } \boldsymbol{\vartheta} = (\mu_m, \boldsymbol{\eta}^T, \boldsymbol{\alpha}^T)^T \text{ and define the}$$

combined estimating function $\psi_{\boldsymbol{\vartheta}}(\mathbf{Z}) = (\psi_{\mu}, \psi_{\boldsymbol{\eta}}^T, \psi_{\boldsymbol{\alpha}}^T)^T$. Then, the estimator

$\hat{\boldsymbol{\vartheta}} = (\hat{\mu}_{\text{DP}}^{(m)}, \hat{\boldsymbol{\eta}}^T, \hat{\boldsymbol{\alpha}}^T)^T$ is the solution to the system of equations, $0 = \mathbb{P}_n\{\boldsymbol{\psi}_{\boldsymbol{\vartheta}}(\mathbf{Z})\}$. Under regularity conditions, $\hat{\boldsymbol{\vartheta}}$ is a consistent estimator for the true value $\boldsymbol{\vartheta}_0$ and $n^{1/2}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)$ converges in distribution to a normal random vector with mean zero and covariance $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T$, where $\mathbf{A} = E\{(\boldsymbol{\psi}_{\boldsymbol{\vartheta}}(\mathbf{Z}))\}$ and $\mathbf{B} = E\{\psi_{\boldsymbol{\vartheta}}(\mathbf{Z})\psi_{\boldsymbol{\vartheta}}^T(\mathbf{Z})\}$. The asymptotic covariance for the direct estimator $\hat{\mu}_{\text{DP}}^{(m)}$ is the upper (1, 1) element of $\boldsymbol{\Sigma}$.

Figure 1.

Illustration of observed data. Y denotes the observed outcome in the end; A_j denotes binary treatment decision to stop or continue at t_j ; X_j denotes covariate information at t_j ; C_j denotes whether the potential treatment terminating event occurs in the interval $(t_j, t_{j+1}]$.

Table 1

Summary statistics for parameter vectors $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ from intermediate statistical models of decision-terminating event and mortality, respectively. Coefficient estimates are pooled estimates from five imputation datasets and standard errors are computed using Rubin's rules.

	Est. (SE)	z-score	p-value
<i>Decision-terminating event</i>			
$\mathbb{I}_{t_0} \quad U < t_1$	-1.26 (0.43)	-2.95	0.003
$\mathbb{I}_{t_1} \quad U < t_2$	-1.79 (0.42)	-4.22	<0.001
$\mathbb{I}_{t_2} \quad U < t_3$	-1.33 (0.41)	-3.21	0.001
$\mathbb{I}_{t_3} \quad U$	-0.35 (0.41)	-0.86	0.387
sex	0.17 (0.13)	1.34	0.180
site of onset	0.87 (0.16)	-5.47	<0.001
age	0.03 (0.01)	4.86	<0.001
current FVC	-0.02 (<0.01)	-5.98	<0.001
<i>4-year mortality event</i>			
Intercept	0.89 (0.92)	0.96	0.336
	0.51 (0.33)	1.54	0.123
$\log(U)$	-0.98 (0.17)	-5.69	<0.001
sex	-0.47 (0.31)	-1.48	0.138
site of onset	-0.32 (0.39)	-0.82	0.413
age	0.05 (0.01)	4.10	<0.001
current FVC	-0.03 (0.01)	-3.71	<0.001

, presence of decision-terminating event; U , observed decision window; sex, female gender is reference; site of onset, spinal site is reference; FVC, percent predicted forced vital capacity

Table 2

Estimated 4-year mortality by PEG insertion policy using data from the Emory ALS Clinic for target insertion times (t_m) after date of diagnosis. Table entries include direct estimates via Monte Carlo (MC) integration and direct prediction (DP), inverse probability weighting (IPW) assuming time-dependent (TD) and baseline (BL) confounding only. The naive estimate is the sample proportion of 4-year mortality for patients belonged to the interval corresponding to t_m . Standard errors are listed in parenthesis following their corresponding estimates.

$t_m(\text{mos.})$	Direct		IPW			Naive
	DP	MC	TD	BL		
3	0.960 (0.016)	0.962 (0.016)	0.984 (0.005)	0.984 (0.006)		0.986 (0.010)
9	0.919 (0.018)	0.927 (0.018)	0.950 (0.021)	0.951 (0.020)		0.943 (0.028)
18	0.888 (0.015)	0.890 (0.015)	0.925 (0.021)	0.922 (0.021)		0.871 (0.043)
36	0.855 (0.020)	0.847 (0.021)	0.868 (0.029)	0.860 (0.025)		0.600 (0.090)

Table 3

Simulation results. Decision-terminating event rate is 40% and 65% in Scenarios 1 and 2, respectively.

n	Method	t _m	Scenario 1				Scenario 2				
			Bias	SSE	SEE	ECP	Bias	SSE	SEE	ECP	
300	Direct	MC	t ₁	-0.003	0.156	0.154	0.947	-0.001	0.165	0.164	0.951
		t ₂	0.001	0.111	0.109	0.937	-0.002	0.119	0.122	0.954	
		t ₃	-0.002	0.103	0.105	0.955	-0.001	0.107	0.111	0.955	
		t ₄	-0.001	0.112	0.122	0.965	-0.004	0.119	0.122	0.957	
		DP	t ₁	-0.001	0.154	0.154	0.947	-0.001	0.156	0.157	0.947
			t ₂	0.001	0.108	0.109	0.951	-0.001	0.121	0.122	0.943
			t ₃	0.004	0.100	0.103	0.950	-0.002	0.105	0.107	0.958
			t ₄	-0.004	0.109	0.115	0.957	-0.003	0.117	0.120	0.956
	IPW	TD	t ₁	0.038	0.409	0.365	0.890	0.073	0.519	0.405	0.829
			t ₂	0.056	0.391	0.328	0.860	0.089	0.482	0.362	0.803
			t ₃	0.035	0.379	0.289	0.839	0.073	0.462	0.309	0.838
			t ₄	-0.002	0.116	0.119	0.950	-0.005	0.103	0.107	0.960
		BL	t ₁	0.341	0.301	0.314	0.780	0.360	0.382	0.367	0.778
			t ₂	0.422	0.255	0.261	0.612	0.380	0.375	0.364	0.751
			t ₃	0.344	0.241	0.234	0.652	0.288	0.362	0.328	0.808
			t ₄	-0.072	0.110	0.113	0.905	-0.047	0.106	0.105	0.928
600	Direct	MC	t ₁	-0.001	0.117	0.115	0.945	-0.008	0.118	0.120	0.949
		t ₂	-0.004	0.087	0.087	0.939	-0.009	0.090	0.091	0.945	
		t ₃	-0.003	0.084	0.083	0.944	-0.006	0.085	0.084	0.944	
		t ₄	-0.003	0.088	0.090	0.948	-0.005	0.093	0.092	0.944	
		DP	t ₁	-0.002	0.111	0.109	0.943	-0.009	0.113	0.115	0.956
			t ₂	-0.003	0.087	0.087	0.942	-0.008	0.089	0.091	0.950
			t ₃	-0.003	0.079	0.079	0.952	-0.006	0.081	0.081	0.951
			t ₄	-0.004	0.090	0.090	0.947	-0.005	0.092	0.091	0.931
IPW	TD	t ₁	0.024	0.303	0.282	0.896	0.009	0.353	0.333	0.901	

<i>n</i>	Method	<i>t_m</i>	Scenario 1				Scenario 2			
			Bias	SSE	SEE	ECP	Bias	SSE	SEE	ECP
BL		<i>t</i> ₂	0.014	0.295	0.259	0.892	0.047	0.363	0.305	0.846
		<i>t</i> ₃	0.018	0.258	0.224	0.882	0.056	0.332	0.258	0.825
		<i>t</i> ₄	−0.002	0.077	0.085	0.966	−0.004	0.075	0.076	0.953
		<i>t</i> ₁	0.354	0.220	0.224	0.623	0.323	0.248	0.270	0.759
		<i>t</i> ₂	0.387	0.183	0.216	0.563	0.376	0.243	0.264	0.683
		<i>t</i> ₃	0.296	0.171	0.173	0.566	0.300	0.257	0.249	0.723
		<i>t</i> ₄	−0.111	0.077	0.075	0.672	−0.047	0.077	0.074	0.906

Bias, Monte Carlo bias; SSE, simulation standard error; SEE, standard error estimate; ECP, empirical coverage probability using Wald-type 95% confidence intervals; MC, direct estimation via Monte Carlo integration; DP, direct estimation via direct prediction; IPW, inverse probability weighted estimator; TD, IPW estimator adjusting for time-dependent factors; BL, IPW estimator adjusting for baseline factors only