Estimating treatment effects from a randomized clinical trial in the presence of a secondary treatment

MIN ZHANG
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA
mzhangst@umich.edu

YANPING WANG
Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

SUMMARY
In randomized clinical trials involving survival time, a challenge that arises frequently, for example, in cancer studies (Manegold, Symanowski, Gatzemeier, Reck, von Pawel, Kortsik, Nackaerts, Lianes and Vogelzang, 2005. Second-line (post-study) chemotherapy received by patients treated in the phase III trial of pemetrexed plus cisplatin versus cisplatin alone in malignant pleural mesothelioma. Annals of Oncology 16, 923-927), is that subjects may initiate secondary treatments during the follow-up. The marginal structural Cox model and the method of inverse probability of treatment weighting (IPTW) have been proposed, originally for observational studies, to make causal inference on time-dependent treatments. In this paper, we adopt the marginal structural Cox model and propose an inferential method that improves the efficiency of the usual IPTW method by tailoring it to the setting of randomized clinical trials. The improvement in efficiency does not depend on any additional assumptions other than those required by the IPTW method, which is achieved by exploiting the knowledge that the study treatment is independent of baseline covariates due to randomization. The finite-sample performance of the proposed method is demonstrated via simulations and by application to data from a cancer clinical trial.

Keywords: Causal inference; Covariate adjustment; Inverse probability weighting; Marginal structural model.

1. INTRODUCTION
Randomized clinical trials are one of the most important tools in determining the causal effect of interventions. In practice, however, despite the careful design of a study, complications may arise in its implementation posing challenges to analysis. This article is motivated by a randomized phase III clinical trial comparing Cisplatin (control) and Pemetrexed/Cisplatin combination in patients with malignant pleural mesothelioma (Vogelzang and others, 2003), a rare tumor linked to asbestos exposure, in terms of the time to all-cause death. It was found that many patients initiated a second-line chemotherapy during the follow-up after discontinuing from the study treatment. More patients in the Cisplatin arm received a secondary treatment, and the time from discontinuation of the study treatment to initiation of a secondary treatment...
treatment was considerably shorter. As the secondary treatment was expected to prolong patients’ survival, naturally, investigators are concerned that such imbalance might have attenuated the “true” effect of the combination treatment. Therefore, it is of interest to estimate the effect of the combination therapy relative to the control under the hypothetical situation that secondary treatments were never used or were balanced between the 2 groups. Such an effect is referred to as the direct effect in epidemiology literature (Rosenblum and others, 2009).

As the decision on whether and when to start a secondary treatment is not part of the study design but rather choices of the patients or their caregivers, such decisions may depend on patients’ baseline and/or postrandomization time-dependent characteristics, which could potentially confound the effect of the study treatments. In the presence of time-dependent confounders, the usual approaches such as the landmark analysis, time-dependent Cox models, and naively censoring subjects at the start of a secondary treatment without further appropriate adjustment do not yield valid inference on the causal effect of treatments. The marginal structural Cox model (Hernán and others, 2000, 2001; Robins and others, 2000) and the method based on inverse probability of treatment weighting (IPTW) have originally been developed for making causal inference from observational studies with time-dependent treatments. Although our setting is a randomized clinical trial, as the secondary treatment is observational in nature, in this article, we adopt the marginal structural Cox model for inference. An alternative to the marginal structural model is the structural nested accelerated failure time model with the inferential method of g-estimation (Hernán and others, 2005). Another alternative is to censor subjects at the initiation of a secondary treatment and then adjust for the induced dependent censoring by inverse probability of censoring weighting (IPCW) (Robins and Greenland, 1994; Zhang and others, 2011), which is discussed in Section 6. Recently, the targeted maximum likelihood method was also proposed to make causal inference for marginal structural models (van der Laan, 2010).

Since secondary treatments are usually not taken into account in determining the required sample size to achieve a certain power, it is critical to use more efficient method in analysis. In this article, we propose a more efficient inferential method for the marginal structural Cox model in the setting of a randomized trial with a secondary treatment. Efficiency gain is achieved in a robust way, in the sense that no additional assumptions other than those required by the usual IPTW method are needed. It has long been recognized that adjusting for covariates leads to increased efficiency of inference in linear regression models (Senn, 1989; Leon and others, 2003). However, in nonlinear models, direct adjustment for covariates may lead to surprising results; for example, Robinson and Jewell (1991) noted that adjusting for covariates in a logistic regression model yields less precise estimators for treatment effects than an unadjusted estimator. This seemingly strange result lies in the fact that in a regression model directly adjusting for covariates, inference is on the treatment effect conditional on covariates, which is a different quantity than the marginal one in a nonlinear regression model, for example, logistic or Cox model (Tsiatis and others, 2007, Section 2). As in many clinical trials, the primary interest is in the marginal treatment effects, in this article, we focus on the marginal treatment effects and, following the principle of Zhang and others (2008) and Lu and Tsiatis (2008), our proposed method improves efficiency by exploiting the knowledge that randomized treatments are independent of baseline covariates.

The remainder of the article is organized as follows. Section 2 sets up the notation and Section 3 describes the proposed method and related background. Simulation studies and analysis of the motivating study are reported in Sections 4 and 5, respectively. The article concludes with a discussion in Section 6.

2. Notation and assumptions

Let Z denote the randomized treatment (Z = 0, 1) and X the vector of baseline covariates, which is assumed to be independent of Z due to randomization. Extension to trials with more than 2 arms is
Treatment effects in the presence of second-line treatment

627

discussed in Section 6. Survival time and censoring time are denoted by \( T \) and \( C \), respectively, with the observed time \( U = \min(T, C) \) and the censoring indicator \( \Delta = (T \leq C) \).

To formally define the parameter of interest, we use the counterfactual outcome framework studied by Rubin (1974). Let \( T^Z \) denote the counterfactual survival time of a subject if, possibly contrary to fact, s/he received the study treatment \( z \) and started a secondary treatment at time \( p \). Corresponding to each subject, conceptually, there is an infinite-dimensional counterfactual since hypothetically a secondary treatment can start at any time. We make the stable unit treatment value assumption (Rubin, 1980), that is, the conceptually, there is an infinite-dimensional counterfactual since hypothetically a secondary treatment

We assume the marginal structural Cox proportional hazards model, where both the randomized and secondary treatments are assumed to affect hazard proportionally, that is,

\[
\lambda_{T^Z}(t) = \lambda_0(t) \exp(\theta_1 z + \theta_2 a(t)),
\]

where \( a(t) \) is an indicator of the secondary treatment with \( a(t) = I(t \geq p) \), \( \lambda_{T^Z}(t) \) is the hazard function of \( T^Z \), and \( \lambda_0(t) \) is the unspecified baseline hazard function. Model (2.1) is simply a Cox model defined in terms of counterfactuals. Under the ideal situation where no one in the study initiated a secondary treatment, \( a(t) = 0 \), and model (2.1) reduces the the usual Cox model in the primary analysis. Therefore, inference based on model (2.1) in the presence of secondary treatments targets the same parameter as in the primary analysis under the ideal situation where there were no secondary treatments.

We define the rest of the notation here. We denote the set of time-dependent covariates at \( t \) by \( V(t) \). The observed data can be summarized as \( O_i = \{Z_i, X_i, U_i, U_i^P, \Delta_i, \Delta_i^P, V_i(t), 0 < t \leq U_i \} \), assumed to be independent and identically distributed across \( i = 1, \ldots, n \), where \( U_i^P = \min(P_i, U_i) \), \( \Delta_i^P = I(P_i < U_i) \). We also define the following notation: \( \theta = (\theta_1, \theta_2) \); \( N_i(t) = I(U_i \leq t, \Delta_i = 1) \) and \( Y_i(t) = I(U_i \geq t) \), which are the observed counting process of death and at-risk process, respectively; and \( N_i^P(t) = I(U_i^P \leq t, \Delta_i^P = 1)I(Z_i = j) \) and \( Y_i^P(t) = I(Y_i(t)I(Z_i = j), j = 0, 1 \).

With regard to the secondary treatment and censoring, we make the usual “no unmeasured confounders” (NUC) assumptions (Hernán and others, 2000, 2001; Robins and Finkelstein, 2000). Specifically, the assumption for \( P \) is given by

\[
\lim h^{-1} P\{t \leq U_i^P < t + h, \Delta_i^P = 1|U_i^P \geq t, Z_i = z, X_i, V_i(u), u \leq t, T_i^Z, p > 0\}
= \lim h^{-1} P\{t \leq U_i^P < t + h, \Delta_i^P = 1|U_i^P \geq t, Z_i = z, X_i, V_i(u), u \leq t\},
\]

which states that, conditional on the randomized treatment, baseline, and time-dependent covariates up to time \( t \), the hazard of initiating a secondary treatment at \( t \) is independent of the potential lifetime of a subject. Similarly with respect to \( C \), the NUC assumption assumes that, conditional on treatment and covariate history up to time \( t \), the hazard of censoring at \( t \) no longer depends on the potential lifetimes.

3. METHODS

In Hernán and others (2000), the IPTW method has been proposed to make inference on model (2.1). Let us first briefly review this method, which we will use as a basis for improvement in efficiency in our proposed method.
3.1 IPTW method

Under the ideal situation where both Z and P are randomized and censoring is independent, inference on a Cox model is routinely made by the maximum partial likelihood (PL) method. In reality, however, the time to secondary treatment may depend on subjects’ past treatment and covariate history. Consequently, the maximum PL method does not make valid inference on model (2.1). The idea of the IPTW method is to inversely weight the contribution of each individual at each risk set by the probability of having the observed history of secondary treatment and the probability of remaining uncensored. Specifically, the IPTW estimator is the solution to the following weighted estimating equation,

$$
\sum_{i=1}^{n} \int [ (Z_i, A_i(t))^T - [Z, A](t; \theta) ] w_i(t) \, dN_i(t) = 0, \tag{3.1}
$$

where $$A_i(t) = I(P_i \leq t)$$, $$[Z, A](t; \theta) = \sum_{j=1}^{n} [(Z_j, A_j(t))^T \exp(\theta_j Z_j + \theta_2 A_j(t)) w_j(t) Y_j(t)]$$, and $$w_i(t)$$ is a weight function, defined later. Note, (3.1) reduces to the usual maximum PL estimating equation if $$w_i(t)$$ is set to 1. The weight $$w_i(t)$$ includes 2 parts, $$w_P(t)$$ and $$w_C(t)$$, corresponding to P and C, respectively. Specifically, $$w_i(t) = w_P(t) w_C(t)$$ with $$w_P(t) = \kappa_P(t; Z_i)/W_P(t; Z_i, X_i, V_i)$$ and $$w_C(t) = \kappa_C(t; Z_i, P_i)/W_C(t; Z_i, P_i, X_i, V_i)$$. The denominator $$W_P(t; Z_i, X_i, V_i)$$ and $$W_C(t; Z_i, P_i, X_i, V_i)$$ are (informally) the probability of having the observed history of secondary treatment conditional on $$[Z_i, P_i, X_i, V_i]$$ and the probability of remaining uncensored by $$t$$ given $$[Z_i, P_i, X_i, V_i]$$, respectively. The numerators $$\kappa_P(t; Z_i)$$ and $$\kappa_C(t; Z_i, P_i)$$ are stabilization factors, intended to stabilize the weights; for example, the numerators being 1 leads to the simple IPTW estimator and the method described below leads to the so-called stabilized weights. In practice, $$w_i(t)$$ is unknown and must be substituted by estimates.

To estimate $$W_P(t; Z_i, X_i, V_i)$$, one need build models for P. A natural choice would be to fit a Cox model for P conditional on $$\{Z, X, V\}$$. To allow for more robustness, we may fit a treatment-specific Cox model, that is,

$$
\lambda^{P}_{ij}(t) = \lambda^{P}_{0j}(t) \exp \{ \gamma^{T}_{j} Q_i(t) \},
$$

where $$\lambda^{P}_{ij}(t)$$ is the conditional hazard of $$P_i$$ for subject i and treatment $$j$$, $$j = 0, 1$$; $$\lambda^{P}_{0j}(t)$$ are unspecified treatment-specific baseline hazard functions; and $$Q_i(t) = \{X_i, V_{it}^{T}(t)\}$$, including baseline and time-dependent covariates. Denoting the estimator for $$\gamma_j$$ by $$\hat{\gamma}_j$$, one then can estimate $$W_P(t; Z_i = j, X_i, V_i)$$ by the Kaplan–Meier type estimator, that is,

$$
\hat{W}_P(t; Z_i = j, X_i, V_i) = \prod_{s < t} \{1 - d\hat{\lambda}^{P}_{ij}(s)\} \text{ if } t < U_i^P,
$$

$$
= \prod_{s < t} \{1 - d\hat{\lambda}^{P}_{ij}(s)\} \{d\hat{\lambda}^{P}_{ij}(U_i^P)\}^{\Delta_i^P} \{1 - d\hat{\lambda}^{P}_{ij}(U_i^P)\}^{-\Delta_i^P} \text{ if } t \geq U_i^P,
$$

where $$d\hat{\lambda}^{P}_{ij}(s) = d\hat{\lambda}^{P}_{0j}(s) \exp \{ \hat{\gamma}_{j}^{T} Q_i(s) \}$$; $$d\hat{\lambda}^{P}_{0j}(s)$$ is the increment of the Breslow estimator of the cumulative baseline hazard function $$\Lambda^{P}_{0j}(s) \equiv \int_{0}^{s} \lambda^{P}_{0j}(s) \, ds$$ with $$d\hat{\lambda}^{P}_{0j}(s) = \frac{\sum_{i=1}^{n} dN_{ij}^{P}(s)}{\sum_{i=1}^{n} \exp \{ \hat{\gamma}_{j}^{T} Q_i(s) \} Y_{ij}(s)}$$. As for the stabilizing factor $$\kappa_P(t; Z_i)$$, one may fit a Cox model for P conditional on Z only and $$\tilde{\kappa}_P(t; Z_i)$$ is the estimated probability of having the observed history of P, calculated similarly as for $$\hat{W}_P(t; Z_i, X_i, V_i)$$. 

Similarly for estimating $W_C(t; Z_i, P_i, X_i, V_i)$, one builds models for $C$ conditional on $(Z_i, P_i, X_i, V_i)$. For example, one can fit a treatment-specific Cox model, using obvious notation given by

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp \{a_j^T B_i(t)\},$$

where $B_i(t) = \{X_i, A_i(t), V_i(t)^T\}^T$. Then, $\hat{W}_C(t; Z_i, P_i, X_i, V_i)$ can be calculated as $\prod_{s \leq t} \{1 - d\hat{\Delta}_{ij}^C(s)\}$ for $t \leq U_i$, where $d\hat{\Delta}_{ij}^C(s)$ is the increment of the Breslow estimator. The numerator $\hat{\kappa}_C(t; Z_i, P_i)$ can be calculated similarly, except that one models $C$ conditional on $(Z_i, P_i)$ only.

If the models for $P$ and $C$ are correctly specified in calculating $\hat{W}_P(t; Z_i, X_i, V_i)$ and $\hat{W}_C(t; Z_i, P_i, X_i, V_i)$, the IPTW estimator, denoted by $\hat{\theta}$, is consistent and asymptotically normal under suitable regularity conditions (Hernán and others, 2000; Robins, 1999). The marginal structural models and the IPTW method have been well studied in both theoretical and applied contexts (Hernán and others, 2000, 2001; Robins and others, 2000; van der Laan and Robins, 2003); We refer readers to these references for more detailed discussion.

### 3.2 Augmented IPTW method

Our proposed method improves the efficiency of the IPTW method by tailoring the inferential method to the setting of randomized clinical trials. Specifically, the proposed method makes use of the knowledge that, due to randomization, $Z$ is independent of all baseline covariates $X$. Following the principle of Zhang and others (2008) and Lu and Tsiatis (2008), we start by considering a class of unbiased estimating functions for $\theta$ that are functions of $X$, which include the estimating function (3.1) of the IPTW method as a special case. Therefore, it allows us to use information contained in $X$ to improve efficiency by identifying the optimal estimating function within this class. The resulting estimators consistently estimate $\theta$, as unbiased estimating functions lead to consistent and asymptotically normal estimators (Stefanski and Boos, 2002).

The IPTW estimator is consistent and asymptotically normal because it is the solution to the unbiased estimating (3.1). Building on (3.1), a class of unbiased estimating equations can be constructed. Subtracting a term that equals zero, we may rewrite (3.1) equivalently as

$$\sum_{i=1}^n \int \left( [Z_i, A_i(t)]^T - [\hat{Z}, \hat{A}](t; \theta) \right) w_i(t) dM_i(t; \theta) = 0,$$

(3.2)

where $dM_i(t; \theta) = dN_i(t) - \exp[\theta_1 Z_i + \theta_2 A_i(t)] \lambda_0(t) Y_i(t) dt$, and we shall use this representation as a basis for improvement. The function inside the summation of an estimating equation is referred to as an estimating function. We denote the estimating function of the IPTW estimator, $\int \left( [Z_i, A_i(t)]^T - [\hat{Z}, \hat{A}](t; \theta) \right) w_i(t) dM_i(t; \theta)$, by $e(O_i; \theta)$, which is unbiased in the sense that $E_0 \{e(O_i; \theta)\} = 0$. Zhang and others (2008) proved that, for randomized clinical trials, a class of unbiased estimating functions can be constructed by augmenting a given unbiased estimating function by some function of the randomized treatment and baseline covariates. Specifically for the estimating function in (3.2), the class of augmented estimating equations are given by

$$\sum_{i=1}^n e(O_i; \theta) - (Z_i - \pi) g(X_i) = 0,$$

(3.3)

where the second term $-(Z_i - \pi) g(X_i)$ is referred to as an augmentation term, $\pi$ is the randomization probability $Pr(Z_i = 1)$, and $g(X_i)$ is an arbitrary function of $X_i$. It is easy to see that the augmentation
term has expectation zero by an iterated conditional expectation argument, that is, \( E\{(Z_i \pi)g(X_i)\} = E\{E\{[Z_i \pi]g(X_i)|X_i|\} = E\{g(X_i)E(Z_i \pi)\} = 0 \), where the second equality is due to \( Z \perp \perp X \). As a result, the augmented estimating function, \( e_i(O_i; \theta) - (Z_i \pi)g(X_i) \), has expectation equal to zero, and therefore, (3.3) is a class of unbiased estimating equations for \( \theta \). Solutions to unbiased estimating equations are referred to as M-estimators (Stefanski and Boos, 2002) or Z-estimators as well, which are consistent and asymptotically normal estimators for the parameter of interest under suitable regularity conditions. Therefore, the estimators obtained by solving (3.3) are consistent and asymptotically normal for \( \theta \).

The class of estimating equations in (3.3) includes (3.2), the estimating equation of the IPTW estimator, as a special case; that is, (3.2) corresponds to an equation in (3.3) with \( g \). In the Supplementary Material (available at Biostatistics online), we showed that the optimal \( g(X_i) \), say \( g_{\text{opt}}(X_i) \), is given by

\[
g_{\text{opt}}(X_i) = \frac{1}{\pi(1 - \pi)} E\{(Z_i - \pi)e(O_i; \theta)|X_i\}. \tag{3.4}
\]

The estimator obtained by solving the estimating equation (3.3) with the optimal \( g(X_i) \) is therefore the estimator with the minimum variance in this class and hence is more efficient than the IPTW estimator. The unknown expectation, \( E\{(Z_i - \pi)e(O_i; \theta)|X_i\} \), in \( g_{\text{opt}}(X_i) \) must be substituted by estimate in practice, which suggests modeling \( (Z_i - \pi)e(O_i; \theta) \) given \( X_i \). For modeling the conditional expectation, we adopt the strategy advocated by Zhang and others (2008), which leads to estimators of \( \theta \) that are guaranteed to be of equal or improved efficiency asymptotically, even if the model for \( E\{(Z_i - \pi)e(O_i; \theta)|X_i\} \) is misspecified. Note that \( e(O_i; \theta) \) is 2-dimensional and we denote the \( k \)th component of \( e(O_i; \theta) \) by \( e_k(O_i; \theta), k = 1, 2 \). We recommend positing a parametric regression model for the conditional expectation, that is,

\[
E\{(Z_i - \pi)e_k(O_i; \theta)|X_i\} = b_k^T q_k(X_i), \tag{3.5}
\]

where \( q_k(X_i) \) are vectors of basis functions in \( X_i \) that may include polynomial terms in elements of \( X \), interaction terms, splines, and so on, and obtaining the estimate of \( b_k \) via the ordinary least squares (OLS) method. Written explicitly, we estimate \( b_k \) by

\[
\hat{b}_k = \left[ \sum_{i=1}^{n}\{q(X_i)q(X_i)^T\} \right]^{-1} \sum_{i=1}^{n}\{q(X_i)(Z_i - \pi)e_k(O_i; \theta)\}. \tag{3.6}
\]

We make predictions for \( g(X_i) \) based on the fitted model by \( \hat{g}(X_i) = \frac{1}{\pi(1 - \pi)}\{\hat{b}_1^T q_1(X_i), \hat{b}_2^T q_2(X_i)\}^T \). Substituting \( \hat{g}(X_i) \) to (3.3), the solution to the estimating equation is the proposed augmented estimator for \( \theta \), which we denote by \( \hat{\theta} \).

We make the following comments about the proposed method. First, as \( e(O_i; \theta) \) depends on the unknown \( \theta \), we must substitute it with an initial estimate in modeling \( E\{(Z_i - \pi)e(O_i; \theta)|X_i\} \), and we suggest replacing it with the usual IPTW estimator \( \hat{\theta} \); similarly, we substitute \( \pi \) with \( \hat{\pi} = \sum_{i=1}^{n} Z_i / n \). As explained in Zhang and others (2008), by the theory of semiparametrics, the asymptotic variance of the estimator obtained by using \( \hat{b}_k \) is the same as that if the limit of \( \hat{b}_k \) is known. As a result, using an inefficient initial estimator for \( \theta \) in the estimation of \( b_k \) will not have an effect asymptotically. Second,
Treatment effects in the presence of second-line treatment

4. Simulation Studies

We carried out simulation studies to evaluate the finite-sample performance of the proposed method, each based on 1000 Monte Carlo data sets with \( n = 400 \). Data were generated under 2 scenarios where both the randomized and secondary treatments have beneficial effect or no effect. In each scenario, we considered 2 cases where the baseline covariates are weakly or strongly associated with the lifetimes. Details on the data generating process are described in the Supplementary Material (available at *Biostatistics* online). We showed that the augmented estimator is guaranteed to be at least as efficient as the unaugmented one if one fits \( b_T q_k(X_i) \) via OLS. Last, variance of the proposed estimator can be estimated by the usual sandwich estimator given by \( \hat{D} = \int \left( \sum_{j=1}^{n} \frac{[Z_i A_i(t)] \exp[ \hat{\theta}_1 Z_i + \hat{\theta}_2 A_i(t)] Y_j(t)}{\sum_{j=1}^{n} \exp[ \hat{\theta}_1 Z_i + \hat{\theta}_2 A_i(t)] Y_j(t)} - \left[ Z, A \right](t; \hat{\theta}) \right)^2 dN_i(t) \) and \( \hat{\Gamma} = 1/n \sum_{i=1}^{n} e(O_i; \hat{\theta}) - (Z_i - \pi) \hat{g}(X_i) \right)^2 \).

We estimate \( \theta \) using the IPTW and the proposed augmented methods, both with the simple and stabilized weights. Results are reported in Tables 1 and 2. All estimators are approximately unbiased and the 95% confidence intervals achieve the nominal level. For estimating \( \theta_1 \), the main quantity of interest, the augmented estimators are more efficient than the corresponding unaugmented ones and, depending on the predictive power of baseline covariates, the augmented method can achieve moderate to considerable gain in efficiency if the stabilized weights are used. Methods using the simple weights are much less efficient than those using the stabilized weights; moreover, the proposed method only achieve very mild efficiency gain over the usual IPTW method using the simple weights. An explanation for these results is that the simple weights are very unstable with some subjects receiving very large weights at some risk sets, whereas the stabilization factors can correct these large weights, making the stabilized weights fairly stable. For example, in our simulated data for the scenario where the association is strong and the treatment effects are nonzero, the interquartile range (IQR) of the simple weights, across subjects and time, is around 1.4 but the standard deviation (SD) is about 85 due to many huge weights; however, for the stabilized weights, both the IQR and SD are much smaller, about 0.11 and 0.25, respectively. Similar patterns of the weights are observed for other scenarios. Although the augmentation is able to improve efficiency theoretically, the effect is mild in the presence of huge weights. For estimating \( \theta_2 \), the augmented estimators are...
### Table 1. Simulation results for estimation of $\theta_1$ and $\theta_2$ in (2.1) in the first scenario

<table>
<thead>
<tr>
<th>Method</th>
<th>Bias</th>
<th>MC SD</th>
<th>Ave. SE</th>
<th>CP</th>
<th>RE</th>
<th>Bias</th>
<th>MC SD</th>
<th>Ave. SE</th>
<th>CP</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weak association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple IPTW</td>
<td>0.008</td>
<td>0.250</td>
<td>0.241</td>
<td>0.939</td>
<td>1</td>
<td>0.012</td>
<td>0.161</td>
<td>0.170</td>
<td>0.964</td>
<td>1</td>
</tr>
<tr>
<td>Aug IPTW</td>
<td>0.006</td>
<td>0.246</td>
<td>0.236</td>
<td>0.940</td>
<td>1.03</td>
<td>0.011</td>
<td>0.163</td>
<td>0.169</td>
<td>0.962</td>
<td>0.97</td>
</tr>
<tr>
<td>Stb IPTW</td>
<td>0.006</td>
<td>0.113</td>
<td>0.129</td>
<td>0.956</td>
<td>1</td>
<td>0.009</td>
<td>0.138</td>
<td>0.151</td>
<td>0.968</td>
<td>1</td>
</tr>
<tr>
<td>Aug Stb IPTW</td>
<td>0.004</td>
<td>0.113</td>
<td>0.116</td>
<td>0.956</td>
<td>1.28</td>
<td>0.008</td>
<td>0.140</td>
<td>0.150</td>
<td>0.967</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Strong association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple IPTW</td>
<td>0.004</td>
<td>0.256</td>
<td>0.248</td>
<td>0.948</td>
<td>1</td>
<td>0.018</td>
<td>0.154</td>
<td>0.168</td>
<td>0.974</td>
<td>1</td>
</tr>
<tr>
<td>Aug IPTW</td>
<td>0.002</td>
<td>0.248</td>
<td>0.240</td>
<td>0.944</td>
<td>1.06</td>
<td>0.018</td>
<td>0.157</td>
<td>0.167</td>
<td>0.973</td>
<td>0.97</td>
</tr>
<tr>
<td>Stb IPTW</td>
<td>0.006</td>
<td>0.130</td>
<td>0.129</td>
<td>0.950</td>
<td>1</td>
<td>0.015</td>
<td>0.130</td>
<td>0.150</td>
<td>0.981</td>
<td></td>
</tr>
<tr>
<td>Aug Stb IPTW</td>
<td>0.004</td>
<td>0.106</td>
<td>0.107</td>
<td>0.957</td>
<td>1.53</td>
<td>0.0139</td>
<td>0.131</td>
<td>0.148</td>
<td>0.980</td>
<td>0.98</td>
</tr>
</tbody>
</table>

“Simple IPTW” refers to the simple IPTW estimator; “Aug IPTW” is the proposed augmented simple IPTW; “Stb IPTW” is the stabilized IPTW estimator; “Aug Stb IPTW” is the augmented stabilized IPTW method. “Weak Association” and “Strong Association” refer to the association between the censored survival time and baseline covariate $X_2$. Bias is Monte Carlo bias, MC SD is Monte Carlo standard deviation, Ave. SE is the average of estimated standard errors obtained using the sandwich formula, CP is the MC coverage probability of 95% Wald confidence intervals, and RE is relative efficiency, calculated as the Monte Carlo mean squared error for the usual IPTW estimator divided by that for the indicated estimator.

### Table 2. Simulation results for estimation of $\theta_1$ and $\theta_2$ in (2.1) in the second scenario, where both randomized treatment and secondary treatment have no effect

<table>
<thead>
<tr>
<th>Method</th>
<th>Bias</th>
<th>MC SD</th>
<th>Ave. SE</th>
<th>CP</th>
<th>RE</th>
<th>Bias</th>
<th>MC SD</th>
<th>Ave. SE</th>
<th>CP</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weak association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple IPTW</td>
<td>0.021</td>
<td>0.264</td>
<td>0.249</td>
<td>0.942</td>
<td>1</td>
<td>0.023</td>
<td>0.171</td>
<td>0.173</td>
<td>0.953</td>
<td>1</td>
</tr>
<tr>
<td>Aug IPTW</td>
<td>0.018</td>
<td>0.260</td>
<td>0.244</td>
<td>0.940</td>
<td>1.03</td>
<td>0.021</td>
<td>0.172</td>
<td>0.172</td>
<td>0.944</td>
<td>0.99</td>
</tr>
<tr>
<td>Stb IPTW</td>
<td>0.009</td>
<td>0.127</td>
<td>0.129</td>
<td>0.953</td>
<td>1</td>
<td>0.011</td>
<td>0.141</td>
<td>0.152</td>
<td>0.967</td>
<td>1</td>
</tr>
<tr>
<td>Aug Stb IPTW</td>
<td>0.007</td>
<td>0.112</td>
<td>0.116</td>
<td>0.958</td>
<td>1.28</td>
<td>0.010</td>
<td>0.142</td>
<td>0.152</td>
<td>0.960</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Strong association</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple IPTW</td>
<td>0.026</td>
<td>0.271</td>
<td>0.255</td>
<td>0.944</td>
<td>1</td>
<td>0.029</td>
<td>0.163</td>
<td>0.170</td>
<td>0.957</td>
<td>1</td>
</tr>
<tr>
<td>Aug IPTW</td>
<td>0.024</td>
<td>0.165</td>
<td>0.248</td>
<td>0.935</td>
<td>1.05</td>
<td>0.027</td>
<td>0.164</td>
<td>0.169</td>
<td>0.954</td>
<td>0.99</td>
</tr>
<tr>
<td>Stb IPTW</td>
<td>0.009</td>
<td>0.129</td>
<td>0.128</td>
<td>0.956</td>
<td>1</td>
<td>0.014</td>
<td>0.132</td>
<td>0.152</td>
<td>0.978</td>
<td>1</td>
</tr>
<tr>
<td>Aug Stb IPTW</td>
<td>0.008</td>
<td>0.104</td>
<td>0.106</td>
<td>0.957</td>
<td>1.54</td>
<td>0.013</td>
<td>0.133</td>
<td>0.151</td>
<td>0.972</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Entries are as in Table 1.

Comparable to the corresponding unaugmented ones in terms of efficiency. There is some finite-sample effect in estimating $\theta_2$ and asymptotically there should be no loss of efficiency. We comment that, in the setting that we are interested in, the effect of the secondary treatment is of much less importance and may be viewed as a nuisance. So improving the efficiency of inference on $\theta_2$ is not our main interest.
We report analysis of the motivating study introduced in the beginning. In this study, 448 patients were randomized to Cisplatin (control) or Pemetrexed/Cisplatin combination with equal probabilities (Cisplatin: 222; Combination: 226). The Kaplan–Meier curves are plotted in Figure 1(a) with the median survival time being 9.3 and 12.1 months, respectively, for the Cisplatin and combination arms. The hazard ratio (HR) from an intent-to-treat perspective was 0.77 (log HR: $-0.27$, standard error [SE]: 0.115), showing that the addition of Pemetrexed has a beneficial effect in terms of survival. Manegold and others (2005) reported that 47.3% patients in the Cisplatin arm received a secondary treatment as opposed to 37.2% in the combination arm and, for patients who received a secondary treatment, the median time from discontinuation of the study treatment to initiation of a secondary treatment was 0.7 and 3.3 months, respectively, for the Cisplatin and combination arms (Vogelzang and others, 2003). To further illustrate the imbalance in the use of secondary treatments, Figure 2 plots the Kaplan–Meier curves for time to initiation of a secondary treatment from the last dose of the study drug. The 2 curves were significantly different (log-rank test: $p$-value $= 0.0038$). The censoring rate was 36% for the combination arm and 28% for the Cisplatin arm with the majority being loss-to-follow-up due to unknown reasons and the rest being administrative censoring.

We carried out the analysis using the IPTW and the proposed methods with the stabilized weights (See Supplementary Material, available at Biostatistics online, for details). The log HR of death corresponding to receiving any secondary treatment versus no secondary treatment was $-0.52$ (SE: 0.221; HR = 0.59) and $-0.54$ (SE: 0.227; HR = 0.58), respectively, from the IPTW and augmented methods, giving evidence that a second-line chemotherapy prolongs patients' survival. After accounting for the secondary treatment, the estimated log HR comparing the combination therapy versus Cisplatin is $-0.36$ (SE: 0.153; HR 0.69) and $-0.43$ (SE: 0.140; HR 0.65), respectively, from the IPTW and augmented methods. Estimators for the HR comparing the randomized treatments after accounting for secondary treatments are smaller than the intent-to-treat analysis, which supports investigators’ concern that, due to imbalance in the use of secondary treatments, the intent-to-treat analysis underestimated the true effect of the study treatment. The proposed method improves efficiency of the estimation of $\theta_1$ by 19% relative to the usual IPTW method, with the relative efficiency calculated as the square of the standard error of $\hat{\theta}_1$ over that of $\tilde{\theta}_1$. In addition
to the Kaplan–Meier curves, which describe the actual survival experience of patients, the survival curves of the 2 arms under the hypothetical situation that no one initiated a second-line chemotherapy using the proposed method are plotted in Figure 1(b).

6. DISCUSSION

We adopted the marginal structural Cox model to account for secondary treatments in a randomized clinical trial and proposed a method that improves efficiency of the usual IPTW method. The proposed method does not require additional assumptions other than those that are assumed for the IPTW method. Simulation studies and application demonstrate that, compared with the usual IPTW method, the proposed method may lead to moderate to considerable gain in efficiency in estimating the effect of the study treatment if the stabilized weights are used.

We focused on clinical trials with 2 arms. The proposed method can be extended to trials with more than 2 arms; see Zhang and others (2008). For a trial with $K$-arms, the augmentation term in (3.3) becomes a summation of $K$ terms, that is, $- \sum_{k=1}^{K} \{ I(Z = k) - P(Z = k) \} g_k(X_i)$, where $g_k(X_i)$ is an arbitrary 2-dimensional function of $X_i$ and the optimal $g_k$ would be $E[e(O_i; \theta)|X_i, Z_i = k]$. It is easy to see that when $K = 2$, this augmentation term is equivalent to the form described before.

The marginal structural Cox model (2.1) assumes the effect of secondary treatments on hazard is proportional. Neither the IPTW method nor the proposed method is robust against misspecification of the effect of secondary treatments. An alternative is to censor subjects at the initiation of a secondary treatment and then to use IPCW to account for the induced dependent censoring. As the effect of secondary treatment is not specified, from this perspective, this approach seems preferable. However, when one artificially censors a subject, information on the subject’s following survival experience is lost, and one might expect this approach less efficient, especially if the percentage of subjects who initiated secondary treatments is high and if the initiation is early in time. Both approaches require correct modeling of the time to secondary treatment on confounders; if relative to this, misspecification of model (2.1) is less of a concern, then methods based on the marginal structural model might be a reasonable strategy. In
Treatment effects in the presence of second-line treatment

(2.1), we assume there is no interaction between the randomized and secondary treatments; however, depending on the application, if it is believed such an interaction exists, the marginal structural model can be extended to accommodate this. Again as this model allows one to exploit subjects’ postsecondary treatment information on survival, benefit in efficiency relative to the method based on artificial censoring can be expected. We point out that the same type of augmentation can be applied to the approach based on the IPCW method to improve efficiency as well.

The augmentation we considered is different from the usual augmented IPTW estimators in that our augmentation term is a function of the randomized treatment, whereas the usual one is a function of the nonrandomized treatment, that is, the treatment used in the weighing (Lunceford and Davidian, 2004). For our problem, the usual augmentation would involve a function of the secondary treatment and the baseline and time-dependent covariates. Informally, our augmentation term is useful to recover information lost by the assignment of $Z$ in the sense that, if a subject is randomized to receive $Z = 0$, his/her potential outcomes under $Z = 1$ are missing, and the usual augmentation is useful to recover information with respect to the secondary treatment. We considered the augmentation as in (3.3) for its simplicity; derivation and implementation of the usual augmentation term for our model are complicated as it involves time-dependent covariates and treatments. It is not clear which augmentation will lead to more efficiency gain; however, we conjecture that the usual augmentation will be more useful to improve the efficiency of estimating $\theta_2$ than $\theta_1$ as it extracts information with respect to the secondary treatment. Greater efficiency gain can be achieved if both augmentation terms are used, which would be interesting for future research.

**SUPPLEMENTARY MATERIAL**

Supplementary material is available at http://biostatistics.oxfordjournals.org.

**ACKNOWLEDGMENTS**

The authors acknowledge Jeremy Taylor, Steve Ruberg, and Ilya Lipkovich for helpful comments. Conflict of Interest: None declared.

**FUNDING**

This research was supported by a grant from the Eli Lilly Corporation.

**REFERENCES**


[Received May 13, 2011; revised March 2, 2012; accepted for publication March 8, 2012]