Published in final edited form as:

Biometrics. 2017 December ; 73(4): 1169–1178. doi:10.1111/biom.12683.

Augmented Estimation for t-year Survival with Censored Regression Models

Yu Zheng* and Tianxi Cai*

*Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

Summary

Reliable and accurate risk prediction is fundamental for successful management of clinical conditions. Estimating comprehensive risk prediction models precisely, however, is a difficult task, especially when the outcome of interest is time to a rare event and the number of candidate predictors, p , is not very small. Another challenge in developing accurate risk models arises from potential model misspecification. Time-specific generalized linear models estimated with inverse censoring probability weighting are robust to model misspecification, but may be inefficient in the rare event setting. To improve the efficiency of such robust estimation procedures, various augmentation methods have been proposed in the literature. These procedures can also leverage auxiliary variables such as intermediate outcomes that are predictive of event risk. However, most existing methods do not perform well in the rare event setting, especially when p is not small. In this paper, we propose a two-step, imputation-based augmentation procedure that can improve estimation efficiency and that is robust to model misspecification. We also develop regularized augmentation procedures for settings where p is not small, along with procedures to improve the estimation of individualized treatment effect in risk reduction. Numerical studies suggest that our proposed methods substantially outperform existing methods in efficiency gains. The proposed methods are applied to an AIDS clinical trial for treating HIV-infected patients.

Keywords

Efficiency augmentation; Intermediate outcomes; Model misspecification; Risk prediction; Robustness; Survival

1. Introduction

Accurate and reliable risk prediction is fundamental for successful disease management, enabling those at high risk of early disease onset to be monitored more closely, and more aggressive treatment options to be considered for patients with poor prognoses. Precise estimation of risk prediction models in the survival setting, however, is difficult especially when the outcome is rare and the number of candidate predictors, p , is not very small. For example, long-term treatment-related co-morbidities, such as Tenofovir-associated renal

Supplementary materials

Web Appendices and Tables referenced in Sections 2 and 3.4 are available with this paper at the *Biometrics* website on Wiley Online Library.

disease in HIV-infected patients, have received increasing attention recently. When initiating a new therapy for HIV-infected patients, it would be of great value to accurately predict the likelihood of patients experiencing long-term treatment-related co-morbidities. However, obtaining a precise risk estimate for these co-morbidities based on clinical studies is challenging due to the limited number of observed events during follow up. Another challenge in developing risk models arises from potential model misspecification. For example, under the commonly used Cox proportional hazards (PH) model (Cox, 1972), regression coefficients can be interpreted as relative risk measures and can reflect the prognostic potential of predictors. With the unknown parameters in the model estimated, it is possible to predict subject-specific survival probabilities for risk classification. However, the PH assumption may not hold in many applications, and prediction models derived under model misspecification may not be validated in future studies since model estimates depend on censoring distributions (Cai and Cheng, 2008; Van Houwelingen, 2007), which are bound to differ across studies.

These challenges signify the importance of developing robust and efficient approaches to the estimation of risk prediction models. Here, we are specifically interested in predicting the risk of a clinical event occurring by time τ using baseline covariates through a τ -year time-specific generalized linear model (τ -GLM), as previously considered in Uno et al. (2007). In the presence of independent censoring and possible model misspecification, Uno et al. (2007) showed that τ -GLM can be estimated consistently using inverse probability weighting (IPW). A risk prediction model is said to be estimated consistently if the model estimates converge in probability to the solution of a limiting estimating equation that is censoring-free. However, such a simple IPW estimator is not efficient as it only uses information from subjects who are not censored prior to τ . Incorporating information from censored subjects could potentially improve estimation efficiency. Additionally, auxiliary variables, including post-baseline intermediate outcomes that are predictive of the primary outcome, may be useful to further improve efficiency (Robins and Rotnitzky, 1992; Gray, 1994; Lu and Tsiatis, 2008; Parast et al., 2014).

There is a rich literature on improving estimation efficiency within an IPW framework by making use of information on auxiliary variables. However, most existing work focuses on non-censored data and simple population parameters such as the mean. Robins et al. (1994) propose a general approach for efficiency augmentation by estimating the selection probability as a function of covariates and extraneous surrogates. However, such a method suffers from the curse of dimensionality (Robins and Ritov, 1997), which would be amplified in the rare event setting. To improve robustness, augmented IPW (AIPW) approaches have been proposed for analyzing survival data (Scharfstein et al., 1999; Bang and Tsiatis, 2000; Tsiatis, 2006; Robins and Wang, 2000), including the τ -GLM (DiRienzo, 2009). The AIPW estimator, involving two separate models that estimate IPW weights and impute outcome, is doubly robust in the sense that it is asymptotically consistent and normally distributed as long as one of the two models is correctly specified. However, the efficiency gain from the AIPW method is often limited by misspecification of the imputation model (Robins et al., 2007). Furthermore, none of the existing methods work well when p is not small.

To overcome the challenges of model misspecification and the curse of dimensionality, we consider a two-step, imputation-based augmentation procedure to improve the estimation of the τ -GLM. In step I, we use a standard IPW approach to approximate the conditional risk function given baseline and surrogate variables under a flexible imputation model. In step II, we impute the τ -year event status based on the estimated conditional risk for each subject, and fit the regression model with imputed outcome using all subjects. Throughout, the risk function of interest is the τ -year risk given the baseline predictors, although post-baseline variables may be used for efficiency augmentation. We demonstrate that with properly chosen imputation procedures, our proposed augmented estimator is consistent under possible misspecification of both the τ -GLM and the imputation model. To further improve efficiency, we construct a combined estimator that is an optimal linear combination of the standard IPW estimator and the augmented estimator. Regularized augmentation procedures are also developed to allow p to remain moderate relative to the number of observed events.

The rest of the paper is organized as follows: in Section 2, we describe the proposed estimation procedure for both risk prediction and the extension to individualized treatment selection settings, where interest lies in estimating subject-specific treatment effects in risk reduction. In Section 3, we present results from simulation studies that demonstrate substantial potential gains in efficiency of estimation under a variety of settings. We illustrate our proposed procedures in Section 4 using an AIDS Clinical Trial for treating HIV patients. Some concluding remarks are given in Section 5.

2. Estimation

Let $T^\dagger$ be a continuous failure time, $\mathbf{X}$ be a $p \times 1$ vector of bounded baseline predictors, and C be the censoring variable. For convenience, we assume that $\mathbf{X} = (1, X_2, \dots, X_p)^\top$. Let $\{(T_i^\dagger, \mathbf{X}_i, C_i), i=1, \dots, n\}$ be n independent copies of $(T^\dagger, \mathbf{X}, C)$. Due to censoring, for the i^{th} subject, we only observe $(T_i, \mathbf{X}_i, \delta_i)$, where $T_i = \min(T_i^\dagger, C_i)$, $\delta_i = I(T_i^\dagger \leq C_i)$ and $I(\cdot)$ is the indicator function. We focus on the setting where $T_i^\dagger$ is subject to heavy censoring, but a $q \times 1$ vector post-baseline outcome, $\mathbf{S}$, is collected before τ and not subject to censoring. We first assume that C is independent of $\{T^\dagger, \mathbf{Z} = (\mathbf{X}^\top, \mathbf{S}^\top)^\top\}$, with a common survival distribution $G(\cdot)$. Extensions to incorporate covariate dependent censoring will be discussed in Web-Appendix A. Let $D = \{(T_i, \delta_i, \mathbf{Z}_i^\top)^\top, i=1, \dots, n\}$ denote the observed data. Throughout, for any vector $\mathbf{a} = (a_1, \dots, a_p)^\top$, $\mathbf{a}_{[-1]} = (a_2, \dots, a_p)^\top$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^\top$.

2.1 Estimation procedure

To predict τ -year survival $Y_\tau = I(T^\dagger \geq \tau)$ based on $\mathbf{X}$, we fit a τ -GLM with logistic link:

$$P(T^\dagger \leq \tau | \mathbf{X}) = P(Y_\tau = 1 | \mathbf{X}) = g(\boldsymbol{\beta}_\tau^\top \mathbf{X}), \quad \text{with } g(x) = e^x / (1 + e^x), \quad (1)$$

where $\boldsymbol{\beta}_\tau$ is the unknown regression parameter. Model (1) allows the predictors to have different effects on long-term and short-term risk, and includes the proportional odds model as a special case. Under (1), Uno et al. (2007) showed that IPW estimator $\tilde{\boldsymbol{\beta}}_\tau$, the solution to

$$\tilde{\mathbf{U}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \hat{w}_i \mathbf{X}_i \{Y_{\tau i} - g(\boldsymbol{\beta}^T \mathbf{X}_i)\}, \quad (2)$$

is a consistent and asymptotically normal estimator of $\bar{\boldsymbol{\beta}}_\tau$, where $\bar{\boldsymbol{\beta}}_\tau$ is the solution to

$$\mathbf{U}_0(\boldsymbol{\beta}) \equiv E[\mathbf{X}\{Y_\tau - g(\boldsymbol{\beta}^T \mathbf{X})\}] = 0,$$

$\hat{w}_i = \{I(T_i \leq \tau)\delta_i + I(T_i > \tau)\}/\hat{G}(T_i \wedge \tau)$ and $\hat{G}(\cdot)$ is the Kaplan-Meier (KM) estimator of $G(\cdot)$. Here, $\bar{\boldsymbol{\beta}}_\tau$ does not depend on the censoring distribution regardless of the adequacy of model (1).

To improve the efficiency of $\bar{\boldsymbol{\beta}}_\tau$ in the presence of model misspecification, we propose a two-step, imputation-based augmentation procedure that aims to make use of all subjects and incorporate additional information on $\mathbf{S}$. In step I, we fit a flexible *working* model,

$P(Y_\tau = 1 | \mathbf{Z}) = g\{\boldsymbol{\theta}_\tau^T \boldsymbol{\Phi}(\mathbf{Z})\}$, to impute Y_τ as $g\{\hat{\boldsymbol{\theta}}_\tau^T \boldsymbol{\Phi}(\mathbf{Z})\}$, where $\boldsymbol{\Phi}(\mathbf{Z})$ is a finite set of basis functions of $\mathbf{Z}$ chosen such that $\mathbf{X}$ is in its linear span. Here, $\hat{\boldsymbol{\theta}}_\tau$ is the minimizer of

$$\hat{R}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \hat{w}_i \ell\{Y_{\tau i}, g(\boldsymbol{\theta}^T \boldsymbol{\Phi}_i)\} + \lambda_n \mathcal{Q}(|\boldsymbol{\theta}_{[-1]}|), \quad \text{with } 0 \leq \lambda_n = o(n^{-\frac{1}{2}})$$

where $\ell(y, x) = y \log\{g(x)\} + (1 - y) \log\{1 - g(x)\}$, $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}(Z_i)$, and $\mathcal{Q}(\cdot)$ is a penalty function such as the ridge or LASSO (Friedman et al., 2001). The tuning parameter λ_n controls the amount of regularization and is chosen via cross validation, but restricted to a specified range to ensure the desired rate. In step II, we estimate $\boldsymbol{\beta}_\tau$ as $\hat{\boldsymbol{\beta}}_\tau$, the solution to

$$\hat{\mathbf{U}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \mathbf{X}_i \left\{ g(\hat{\boldsymbol{\theta}}_\tau^T \boldsymbol{\Phi}_i) - g(\boldsymbol{\beta}^T \mathbf{X}_i) \right\}.$$

In Web-Appendix B, we show that $\hat{\boldsymbol{\beta}}_\tau$ is a consistent estimate of $\bar{\boldsymbol{\beta}}_\tau$ even when $\mathbf{S}$ is measured post-baseline. In addition, we show in Web-Appendix C that

$$\tilde{\mathbb{W}}_\tau \equiv n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\tau - \bar{\boldsymbol{\beta}}_\tau) = n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\eta}_{\text{IPW},i} + o_p(1) \text{ and}$$

$$\hat{\mathbb{W}}_\tau \equiv n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\tau - \bar{\boldsymbol{\beta}}_\tau) = n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\eta}_{\text{AUG},i} + o_p(1) \text{ where}$$

$$\begin{aligned} \boldsymbol{\eta}_{\text{IPW},i} &= \mathbf{F}_{1i} + (w_i - 1)\mathbf{F}_{1i} - \int_0^\tau \psi_i(s) \boldsymbol{\mu}_{F_1}(ds), \quad \mathbf{F}_{1i} = -1 \mathbf{X}_i \{Y_{\tau i} - g(\bar{\boldsymbol{\beta}}_\tau^T \mathbf{X}_i)\} \\ \boldsymbol{\eta}_{\text{AUG},i} &= \mathbf{F}_{1i} + (w_i - 1)\mathbf{F}_{2i} - \int_0^\tau \psi_i(s) \boldsymbol{\mu}_{F_2}(ds), \quad \mathbf{F}_{2i} = -1 \mathbf{X}_i \{Y_{\tau i} - g(\hat{\boldsymbol{\theta}}_\tau^T \boldsymbol{\Phi}_i)\}, \end{aligned}$$

$$w_i = \{I(T_i \leq \tau)\delta_i + I(T_i > \tau)\}/G(T_i \wedge \tau), \quad \psi_i(s) = \int_0^s \pi(u)^{-1} dM_{ci}(u), \quad \pi(u) = P(T_i > u),$$

$$M_{ci}(u) = I(T_i \leq u, \delta_i = 0) - \int_0^u I(T_i > v) d\Lambda_c(u), \quad \Lambda_c(u) = -\log\{G(u)\}, \quad E[\mathbf{X}\mathbf{X}^T \dot{g}(\bar{\boldsymbol{\beta}}_\tau^T \mathbf{X})],$$

and $\boldsymbol{\mu}_{F_k}(s) = E\{\mathbf{F}_{ki}I(T_i^\dagger > s)\} = -E\{\mathbf{F}_{ki}I(T_i^\dagger \leq s)\}$ for $k=1,2$. It follows that $\hat{\mathbb{W}}_\tau$ and $\tilde{\mathbb{W}}_\tau$ converge in distribution to zero-mean multivariate normals with covariance matrices $\sum_{\text{AUG}} = E(\boldsymbol{\eta}_{\text{AUG},i}^{\otimes 2})$ and $\sum_{\text{IPW}} = E(\boldsymbol{\eta}_{\text{IPW},i}^{\otimes 2})$, respectively. It is also shown in Web-Appendix C that the variance reduction $\text{var} = \Sigma_{\text{AUG}} - \Sigma_{\text{IPW}}$ takes the form

$$\int_0^\tau \{\text{var}(\mathbf{F}_{1i}|T_i^\dagger > s) - \text{var}(\mathbf{F}_{2i}|T_i^\dagger > s)\} \frac{S(s)^2 d\Lambda_c(s)}{\pi(s)} + \{\boldsymbol{\mu}_{F_1}(\tau)^{\otimes 2} - \boldsymbol{\mu}_{F_2}(\tau)^{\otimes 2}\} \int_0^\tau \frac{d\Lambda_c(s)}{\pi(s)}.$$

Although $\hat{\beta}_\tau$ is often more efficient than $\tilde{\beta}_\tau$ since the imputation leverages potential non-linear effects and information on $\mathbf{S}$, $\hat{\beta}_\tau$ is not fully efficient, especially when the imputation model $P(Y_\tau=1|\mathbf{Z})=g\{\boldsymbol{\theta}_\tau^\top \boldsymbol{\Phi}(\mathbf{Z})\}$ is misspecified. This motivates us to propose our final estimator as the optimal linear combination of the two candidate estimators, $\tilde{\beta}_\tau$ and $\hat{\beta}_\tau$. For simplicity, we only consider component-wise optimal linear combinations and construct

$\hat{\beta}_{\tau j}^{\text{comb}} = \hat{\mathbf{C}}_j^\top (\tilde{\beta}_j, \hat{\beta}_j)^\top$, where $\hat{\mathbf{C}}_j = \mathbf{1}^\top \sum_j^{-1} / \mathbf{1}^\top \sum_j^{-1} \mathbf{1}$ and $\sum_j$ is a consistent estimator for the covariance matrix of $(\tilde{\mathbb{W}}_{\tau j}, \hat{\mathbb{W}}_{\tau j})^\top$. It is straightforward to see that $\hat{\beta}_{\tau j}^{\text{comb}}$ is always at least as efficient as $\tilde{\beta}_{\tau j}$ or $\hat{\beta}_{\tau j}$. Such a linear combination is desirable because $\tilde{\beta}_{\tau j} - \hat{\beta}_{\tau j}$ is not always independent of $\hat{\beta}_{\tau j}$ and as such, the combined estimator would be more efficient than $\hat{\beta}_{\tau j}$.

2.2 Inference via Resampling

To obtain the combined estimator as well as its corresponding variance, we propose a perturbation resampling procedure similar to those proposed in Uno et al. (2007) for approximating the distribution of $\tilde{\mathbb{W}}_\tau$. Let $\mathbf{V} = (V_1, \dots, V_n)^\top$ be an $n \times 1$ vector of independent and identically distributed random variables with mean 1 and variance 1, generated independent of $\mathcal{D}$. Then we obtain a perturbed version of $\hat{\beta}_\tau$ as $\hat{\beta}_\tau^*$, the solution to

$$\hat{\mathbf{U}}_n^*(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n V_i \mathbf{X}_i \{g(\boldsymbol{\Phi}_i^\top \boldsymbol{\theta}^*) - g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} = 0$$

where $\hat{\boldsymbol{\theta}}_\tau^*$ is the minimizer of $\hat{R}_n^*(\boldsymbol{\theta}) \equiv n^{-1} \sum_{i=1}^n \hat{w}_i^* \ell\{Y_{\tau i}, g(\boldsymbol{\theta}^\top \boldsymbol{\Phi}_i)\} + \lambda_n \mathcal{Q}(|\boldsymbol{\theta}_{[-1]}|)$, $\hat{w}_i^* = V_i \{I(T_i \leq \tau) \delta_i + I(T_i > \tau)\} / \hat{G}^*(T_i \wedge \tau)$,

$$\hat{G}^*(t) = \hat{G}(t) \exp \left\{ -n^{-1} \sum_{i=1}^n \int_0^t \frac{(V_i - 1) d\hat{M}_{ci}(s)}{\hat{\pi}(s)} \right\},$$

$\hat{M}_{ci}(t) = I(T_i \leq t, \delta_i = 0) + \int_0^t I(T_i > s) d\log\{\hat{G}(s)\}$ and $\hat{\pi}(s) = n^{-1} \sum_{i=1}^n I(T_i > s)$. The tuning parameter λ_n can be either reselected for each perturbation or fixed as the initial λ_n selected

for the observed data. The latter approach substantially reduces the computational cost and appears to perform well in our simulation studies. However, when computation is not a concern, we recommend reselecting the optimal tuning parameter for each perturbed sample to ensure the performance of the combined estimator.

In practice, one can generate a large number, say B , of random samples of $\mathbf{V}$ to obtain B realizations of $\tilde{\beta}_\tau^*$ and $\hat{\beta}_\tau^*$. Then the unconditional distribution of $(\tilde{\zeta}_\tau^\top, \hat{\zeta}_\tau^\top)^\top$ can be approximated by the empirical distribution of $\{n^{\frac{1}{2}}(\tilde{\beta}_\tau^* - \tilde{\beta}_\tau), n^{\frac{1}{2}}(\hat{\beta}_\tau^* - \hat{\beta}_\tau)\}^\top$, conditional on $\mathcal{D}$. This joint distribution can be subsequently used to construct $\hat{\beta}_\tau^{\text{comb}}$ and the confidence intervals (CIs) for the combined estimator. However, in the realistic situation where sample size is limited, $\sum_j^{-1}$ could be unstable due to high correlation between $\tilde{\beta}_\tau$ and $\hat{\beta}_\tau$, in which case $\text{var}(\hat{\beta}_j^{\text{comb}})$ is underestimated. One solution is to use $(\sum_j + \varepsilon \mathbf{I})^{-1}$ instead of $\sum_j^{-1}$, where ε is a small number that can be chosen in the order of $O(n^{-\frac{1}{2}})$. For instance, we use $\varepsilon = \{\text{var}(\tilde{\beta}) + \text{var}(\hat{\beta})\} / (2n^{1/2})$ in the example illustrated in Section 4. The standard error (SE) estimates and CIs for various parameters can be constructed empirically based on these resampled realizations. Note that the perturbed scheme is similar to standard bootstrap if $\mathbf{V}$ is generated from a multinomial distribution. However, in the case of rare events, bootstrap is likely to suffer from instability since the subsampling might lead to bootstrap samples with too few events, while the proposed perturbation scheme avoids such a problem by assigning positive weights to all observations.

2.3 Improving Estimation of Individualized Treatment Effect (ITE)

When interest lies in identifying individualized treatment rules (ITRs), we assume that data for analysis arise from randomized clinical trials with two treatment groups indexed by $A \in \{0, 1\}$. We use $Y_\tau^{(a)}$ to denote the counterfactual τ -year survival status if a patient is assigned to treatment $A = a$. The optimal binary ITR $\mathcal{J}(\mathbf{X}): \mathbf{X} \rightarrow \{0, 1\}$ maximizing the population average value function, $(\mathcal{J}) = 1 - E\{\mathcal{J}(\mathbf{X})Y^{(1)} + (1 - \mathcal{J}(\mathbf{X}))Y^{(0)}\}$ is the Bayes rule $I\{D(\mathbf{X}) \geq 0\}$, where $D(\mathbf{X}) = E(Y^{(1)}|\mathbf{X}) - E(Y^{(0)}|\mathbf{X})$ (Matsouaka et al., 2014). Under working models $E(Y_\tau^{(a)}|\mathbf{X}) = g(\beta_{\tau,a}^\top \mathbf{X})$, for any given $\mathbf{x}$ that is a realization of $\mathbf{X}$, we may approximate $D(\mathbf{x})$ by $\bar{D}(\mathbf{x}) = g(\bar{\beta}_{\tau,1}^\top \mathbf{x}) - g(\bar{\beta}_{\tau,0}^\top \mathbf{x})$, where $\bar{\beta}_{\tau,a}$ is the solution to $E[\mathbf{X}\{Y_\tau^{(a)} - g(\beta^\top \mathbf{X})\}] = 0$. Our proposed method could be used to obtain more efficient estimation of $\bar{\beta}_{\tau,a}$ and subsequently a more efficient estimator of $\bar{D}(\mathbf{X})$.

2.4 Incorporating Moderate p

When p is not small relative to the number of events, one may obtain a regularized IPW estimator of β_τ . To this end, we note that $\tilde{\beta}_\tau$ and the augmented estimator $\hat{\beta}_\tau$ are the respective minimizers of $\tilde{L}_n(\beta) = -\sum_{i=1}^n \hat{w}_i \ell(Y_{\tau i}, \beta^\top \mathbf{X}_i)$ and

$\hat{L}_n(\boldsymbol{\beta}) = -\sum_{i=1}^n \ell\{g(\hat{\boldsymbol{\theta}}_\tau^\top \Phi_i), \boldsymbol{\beta}^\top \mathbf{X}_i\}$. Then we estimate $\boldsymbol{\beta}_\tau$ as $\tilde{\boldsymbol{\beta}}_\tau$, the minimizer of the adaptive LASSO (Zhang and Lu, 2007) penalized objective function,

$$\tilde{L}_n(\boldsymbol{\beta}) + \tilde{\nu}_n \sum_{j=2}^p |\beta_j / \tilde{\beta}_{\tau j}|$$

where $\tilde{\nu}_n \geq 0$ controls the amount of regularization and $\tilde{\nu}_n \rightarrow 0$ as $n \rightarrow \infty$. Using similar arguments as given in Zou (2006), one may show that $\tilde{\boldsymbol{\beta}}_\tau \rightarrow \bar{\boldsymbol{\beta}}_\tau$ in probability. Similarly, we

may obtain a regularized version of $\hat{\boldsymbol{\beta}}_\tau$, $\hat{\boldsymbol{\beta}}_{\tau^*}$, as the minimizer of $\hat{L}_n(\boldsymbol{\beta}) + \hat{\nu}_n \sum_{j=2}^p |\beta_j / \hat{\beta}_{\tau j}|$ for some properly chosen tuning parameter $\hat{\nu}_n \geq 0$. In Web-Appendix D, we show that

$n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_\tau - \bar{\boldsymbol{\beta}}_\tau)$ and $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{\tau^*} - \bar{\boldsymbol{\beta}}_\tau)$ converge in distribution to zero-mean multivariate normals. To assess the variability of $\tilde{\boldsymbol{\beta}}_\tau$ and $\hat{\boldsymbol{\beta}}_{\tau^*}$, similar perturbation resampling procedures can be employed.

3. Simulations

We conducted extensive simulation studies with sample size $n = 500$ to examine the finite sample properties of the proposed estimation procedures and compare them to existing methods. For the enriched imputation model, we used natural spline bases with 3 knots for each of the covariates across all settings. For each dataset, we used 500 perturbations for variance estimation. Efficiency gain is defined as the relative mean square error minus one in percentage scale. All results are based on 500 simulated datasets for each configuration. We first considered independent censoring in Sections 3.1–3.3 and generated C from an exponential with mean λ , chosen to achieve designed event rates. In Section 3.4, we investigated the case with C dependent on $\mathbf{X}$, and the robustness of various procedures with respect to estimation and prediction. Note that the effective sample size is much smaller than 500 due to the relatively low event rates.

For comparison, we also obtained (i) IPW estimators with censoring weights calculated from KM estimator (IPW_{KM}) as in Uno et al. (2007); (ii) IPW estimators with censoring weights estimated by fitting a Cox model with both $\mathbf{X}$ and S (IPW_{COX}) as in Scharfstein et al. (1999); (iii) an AIPW estimator constructed as in DiRienzo (2009) with censoring weights estimated by KM (AIPW_{KM}); and (iv) an AIPW approach similar to (iii) with the censoring weights calculated using a Cox model (AIPW_{COX}). For our methods, we were also interested in evaluating the added value of S in improving efficiency, and thus constructed the augmented estimators with the imputation step using (I) $\mathbf{X}$ only (AUG_X); and (II) both $\mathbf{X}$ and S ($\text{AUG}_{X,S}$). Finally, we also considered the proposed combined estimator (AUG_{CMB}) that optimally combines IPW_{KM} and $\text{AUG}_{X,S}$.

3.1 Low Dimensional Baseline Covariates

We first considered a small $p = 4$ and generated $\mathbf{X}_{-1}$, T^* and S from

$$\mathbf{X}_{-1} = (X_2, X_3, X_4)^T \sim N(\mathbf{0}, 0.3 + 0.7\mathbb{I}_3) \text{ where } \mathbb{I}_d \text{ is the } d \times d \text{ diagonal matrix,} \\ (\mathcal{M}_1): \log(T^\dagger) = 0.5(X_2 + X_3 + X_4) + 0.5X_2^2 + X_3^2 + 0.5X_4^2 - 3 + \text{logit}(U) + \log(\alpha), \\ S = \text{logit}(U) + 0.1X_2 + 0.1X_3 + \sigma_s \varepsilon \text{ with } U \sim \text{Uniform}(0, 1) \text{ and } \varepsilon \sim N(0, 1),$$

We let $\sigma_s = 0.5$ or 2 to induce high or moderate correlation between S and $\log(T^\dagger)$, respectively. We generated C from $\mathcal{C}_{\text{KM}} \sim \exp(\lambda)$, an exponential distribution with mean λ . We considered (i) a low event rate (12–18% by τ) and heavy censoring setting (65–74% before τ) with $\{\alpha = 12, \lambda = 0.5\}$; and (ii) a moderate event rate (25–34% by τ) and moderate censoring setting (37–50% before τ) with $\{\alpha = 6, \lambda = 1\}$, where $\tau = 0.8$.

As shown in Table 1, estimators of β_τ have negligible biases across all settings. Our proposed estimators AUG_X , $\text{AUG}_{X,S}$ and AUG_{CMB} perform substantially better than all existing estimators with efficiency gains relative to IPW_{KM} as high as 106%, while existing augmented estimators such as AIPW_{cox} only have very modest gains. In general, $\text{AUG}_{X,S}$ performs better than AUG_X , which indicates that incorporating S can lead to additional efficiency gains. The resampling-based inference procedures also perform well with the SE estimates close to the empirical SE and the empirical coverage probability (CovP) of the 95% CIs close to their nominal level.

3.2 ITE Estimation

We also studied the performance of our proposed estimators for ITE under a randomized clinical trial setting with treatment group $A \in \{0, 1\}$ and $n_a = 500$ for $a = 0, 1$. We generated $\mathbf{X}_{-1} = (X_2, X_3, X_4)^T$ from $N(\mathbf{0}, 0.3 + 0.7\mathbb{I}_3)$ for both groups, and $T^\dagger$ from $\log(T^\dagger) = (0.1 + 0.1A)X_2 + (0.4 - 0.3A)X_3 + (0.1 + 0.2A)X_4 + 0.5X_2^2 + X_3^2 + 0.5X_4^2 - 3 + \text{logit}(U) + \log(12 - 4A)$ with $U \sim \text{Uniform}(0, 1)$. The surrogate variable $S = \text{logit}(U) + 0.5A + 0.1X_2 + 0.1X_3 + 0.3\varepsilon$, where $\varepsilon \sim N(0, 1)$ and the correlation between S and $\log(T^\dagger)$ was around 60% within group. We generated $C|A$ from an exponential distribution with mean 0.6 leading to event rates by τ of about 16–27% for $A = 1$ and 12–21% for $A = 0$. We obtained estimates of $\bar{\beta}_{\tau_a}$ for $\tau = 0.9$ using various methods and then estimated the ITE score $\bar{D}(\mathbf{x}) = g(\bar{\beta}_{\tau,1}^\top \mathbf{x}) - g(\bar{\beta}_{\tau,0}^\top \mathbf{x})$ for two specific covariate levels: (i) $\mathbf{x} = (1, 1, 1, 1)^\top$; and (ii) $\mathbf{x} = (1, -0.5, -0.5, -0.5)^\top$.

In Table 2, we compared the model-based estimate of conditional τ -year risk for patients treated with either $A = 1$ or $A = 0$ and the ITE score $D(\mathbf{x})$ using regression coefficients obtained from various methods. Again, while all estimators have negligible bias, the proposed methods are substantially more efficient than existing methods. The proposed resampling procedures also work well for making inference with proper SE estimation and empirical coverage for the CIs.

3.3 Baseline Model Regularization

We also conducted simulation studies to assess the finite sample performance of the proposed methods that incorporate regularization with $p = 11$. We generated $\mathbf{X}_{-1} \sim N(\mathbf{0}, \mathbb{I}_{10})$ and

$$\log(T^\dagger) = X_2 + X_3 + X_4 + 0.5X_2^2 + X_3^2 + 0.5X_4^2 - 3 + \text{logit}(U) + \log(6).$$

We generated C and S as in Section 3.1 and let $\lambda = 1$, $\sigma_s = 0.5$ and $\tau = 1$, leading to an observed event rate by τ of about 26 – 38%, and a correlation between S and $\log(T^\dagger)$ of about 0.55 – 0.7. The effective sample size is around 100–200, not large relative to $p = 11$. To control overfitting in the imputation model, L_1 penalized logistic regression (Friedman et al., 2010) with IPW was employed when estimating the imputation model.

Figure 1 summarizes the results for our estimators, as well as IPW_{KM} and IPW_{COX} as benchmarks for efficiency assessment. The AIPW methods were not included as no associated regularization procedures were available. The proposed methods that incorporate both $\mathbf{X}$ and S are generally more efficient than IPW_{COX} and IPW_{KM} , as in the previous settings. The resampling procedures also perform well with empirical CovP ranging from 92–95% for informative signals. The CovPs for the zero signals range from 96%–98%, which is expected owing to the oracle properties.

3.4 Dependent censoring

We next investigated the robustness of various estimators with respect to the independent censoring assumption. We focused on $\hat{\beta}_\tau^{\text{comb}}$ and a variation of our estimator with IPW weights estimated from a Cox model, denoted by $\text{AUG}_{\text{CMB}}^{\text{COX}}$. We let $p = 4$, generate $\mathbf{X}_{-1}$, S , $T^\dagger$ from $(\mathcal{M}_1)$ with $\alpha = 12$ and $\mathcal{C}_{\text{KM}} \sim \exp(\lambda)$ with $\lambda = 0.5$, and then generated C from either

$$\begin{aligned} (\mathcal{C}_{\text{COX}}): & \text{a Cox model with } \mathcal{C} = \mathcal{C}_{\text{KM}} \exp(\alpha_1^T \mathbf{X}_{-1}); \text{ or} \\ (\mathcal{C}_{\text{NCOX}}): & \text{a non-Cox model with } \mathcal{C} = \mathcal{C}_{\text{KM}} + \alpha_2^T \exp(-\mathbf{X}_{-1}) \end{aligned}$$

where $\alpha_1 = (-0.2, 0, 0)^T$ and $\alpha_2 = (0.2, 0, 0)^T$. Under such parameterization, censoring is associated with X_2 with a hazard ratio between 1.1 and 1.9. Results are summarized in Web Table 1 for $\tau = 0.8$. When C is from $(\mathcal{C}_{\text{COX}})$, IPW_{COX} , AIPW_{COX} , and $\text{AUG}_{\text{CMB}}^{\text{COX}}$ all have negligible bias with proper coverage levels for the CIs. This is as expected, since these methods calculate IPW weights from fitting a Cox model for C . When C is from $(\mathcal{C}_{\text{NCOX}})$, none of these methods provide consistent estimators, but AUG_{CMB} and $\text{AUG}_{\text{CMB}}^{\text{COX}}$ generally have much smaller bias than existing estimators. The estimator AUG_{CMB} is also not very sensitive to the departure of the independent censoring assumption. Compared to the benchmark estimator of IPW with KM weights, our estimators are substantially more efficient with respect to mean square error.

Although τ -GLM is a more flexible model, its associated inference procedures require stronger assumptions regarding the censoring distribution. The traditional Cox model makes stronger model assumptions but requires weaker censoring assumptions. To compare different procedures including the Cox model under model misspecification, we compared the performance of various procedures with respect to the accuracy in predicting τ -year survival based on: (i) the area under the Receiver Operating Characteristic curve (AUC); and (ii) incremental proportion of explained variance (IPEV), defined as

$1 - E[\{I(T^\dagger \leq \tau) - \hat{\mathcal{P}}_\tau(\mathbf{X})\}^2] - E[\{I(T^\dagger \leq \tau) - \hat{P}_\tau\}^2]$ (Gerds et al., 2008), where $\hat{\mathcal{P}}_\tau(\cdot)$ is the estimated conditional risk function, and $\hat{P}_\tau$ is the KM estimator of $P(T^\dagger \leq \tau)$ using training data. For each trained model, the corresponding accuracy parameters are estimated via Monte Carlo using a fully observed independent validation dataset with a sample size of 100,000. The true underlying data are generated from two misspecified models: $(\mathcal{M}_1)$ as defined previously in this section; and $(\mathcal{M}_2)$ $\mathbf{X}_{-1} = (X_2, X_3)^\top$ generated from multivariate log-normal with mean zero, covariance (0.1, 1.2) and correlation -0.1, $\log(T^\dagger) = \log(X_2^2 + 3X_3^2) + 0.1(X_2^2 + 2X_3^2)\xi + 0.5$ and $S = 0.4\log(X_2^2 + X_3^2)\xi + \varepsilon$ with $\xi \sim \text{logistic}(0, 1)$, $\varepsilon \sim \mathcal{N}(0, 1)$. The τ -GLM misspecifies the covariate effects for Model $(\mathcal{M}_1)$, and does not capture the covariate effects or heteroskedasticity correctly for $(\mathcal{M}_2)$. For both $(\mathcal{M}_1)$ and $(\mathcal{M}_2)$, we considered three different censoring distributions $(\mathcal{C}_{\text{KM}})$, $(\mathcal{C}_{\text{cox}})$ and $(\mathcal{C}_{\text{ncox}})$. For Model $(\mathcal{M}_2)$, we let $\lambda = 1$, $\mathbf{a}_1 = \mathbf{a}_2 = (0.2, 0)^\top$, and $\tau = 0.5$, which is about the 70th percentile of the observed follow-up time. For the imputation step of our methods, we used 5-knot natural spline bases. The results are shown in Web Table 2. They suggest that our proposed procedures are robust in model misspecification and covariate dependent censoring with respect to prediction performance. Compared to the Cox model, the proposed method achieves similar prediction accuracy under $(\mathcal{M}_1)$, and substantially higher accuracy under $(\mathcal{M}_2)$, across different censoring patterns.

4. Example

We illustrate the proposed procedures using a dataset from the AIDS Clinical Trial Group (ACTG) Protocol 175 (Hammer et al., 1996). This study consists of 2467 patients randomized to 4 different treatments; zidovudine (ZDV) only, ZDV+didanosine (ZDV+DDI), ZDV+zalcitabine (ZDV+ZAL), and didanosine (DDI) only. For illustration, we focus on patients treated with ZDV only (mono group, n=619) and ZDV+DDI (combo group, n=612) and use the study-defined secondary endpoint – ‘time to progression’ defined as occurrence of death or AIDS as defining events. Our goal is to develop (I) risk prediction models for progression by week 144 within each treatment group; and also (II) an ITE prediction model. Baseline covariates, $\mathbf{X}$, include baseline CD4 counts (CD4), age, intravenous drug use (IV), use of ZDV within 30 days prior to randomization (pZDV), and symptomatic status (SS). We include three intermediate outcomes for $\mathbf{S}$: (i) short-term immune response defined as CD4 count at week 24; and (ii) short-term toxicity indicating whether the patient has experienced grade 3 or higher abnormal lab results by week 24; and (iii) short-term tolerability indicating whether the patient has experienced grade 3 or higher signs and symptoms by week 24. Missing values were handled by carrying forward the last observation. Note that less than 0.5% subjects experienced an event of interest and another 3.5% subjects were censored before week 24. These subjects were excluded from the analyses.

We first consider risk prediction within each treatment group based on $\mathbf{X}$ using various methods. For the enriched imputation modeling required by our approach, we use spline bases with 3 knots for all continuous variables. Resampling with 500 replications is used to generate the variance of the IPW_{KM} method and our proposed methods, and bootstrap was

used for other methods. As shown in Table 3, the point estimates from different methods are comparable to each other. CD4 is a significant risk predictor for both treatment groups, while SS and pZDV are only significant for those treated by ZDV+DDI. In general, our estimators are substantially more efficient than alternative methods. For the ZDV group, the efficiency gain appears to come primarily from augmentation from $\mathbf{X}$ alone. The intermediate outcomes do not appear to be highly predictive of the progression, and consequently, incorporating $\mathbf{S}$ in the augmentation resulted in moderate efficiency loss due to overfitting. On the other hand, for the ZDV+DDI group, incorporating $\mathbf{S}$ substantially improves the efficiency of the estimators. The combined estimator AUG_{CMB} generally performs well.

Based on these risk models, we also derived models for predicting the individualized treatment benefit of ZDV+DDI versus ZDV alone on the progression risk at week 144. In addition to IPW methods, we also added the prediction results from the Cox model. The predicted risks within each treatment group for two examples of covariates, along with their specific treatment differences, are presented in Table 4. Example 1 is a 22-year old non-IV drug user who had a baseline CD4 count 100 cells/ mm^3 without symptoms at baseline, and did not take ZDV 30 days prior to randomization. The treatment difference for this patient was estimated around -0.18 with SE around 0.12 based on our methods, suggesting a moderate treatment difference. Note that the proposed estimators provide more efficient estimation in treatment difference than all other methods including the Cox model. The Cox model has a smaller SE in risk probability estimation for the ZDV+DDI arm, possibly due to smaller estimated probability. Example 2 is a 40-year old IV drug user who had a baseline CD4 count 350 cells/ mm^3 without symptoms at baseline, and took ZDV 30 days prior to randomization. For this patient, the treatment difference was estimated as -0.09 with SE around 0.05, also suggesting a modest treatment difference. For this patient, all the methods provide similar point estimates and the proposed estimators are substantially more efficient than other methods, including the Cox model.

5. Remarks

Leveraging information on post-baseline intermediate variables and allowing for possible model misspecification, our proposed robust IPW estimating procedure improves the efficiency of covariate-specific t -year survival prediction by making use of information from censored subjects. Compared with other methods, our approach has the following advantages: (i) the consistency of the estimator and the efficiency gain do not rely on the correctly-specified working model, and the degree of efficiency gain is actually even larger under the incorrect specification of the risk model; (ii) by allowing a flexible enriched imputation model that incorporates non-linear effects, the proposed method can achieve higher efficiency gains; and (iii) the incorporation of regularization for settings where p is not small relative to the effective sample size.

The proposed estimator $\hat{\beta}$ is related to the DiRienzo (2009) AIPW estimator, which is the solution to $n^{-1} \sum_{i=1}^n \mathbf{X}_i \{ \hat{Y}_i + \hat{w}_i(Y_{\tau i} - \hat{Y}_i) - g(\beta^\top \mathbf{X}_i) \} = 0$, where $\hat{Y}_i = g(\hat{\gamma}^\top \mathbf{X}_i)$ and $\hat{\gamma}$ is the solution to $0 = n^{-1} \sum_{i=1}^n \{ I(T_i \leq \tau) \delta_i + I(T_i > \tau) \} \mathbf{X}_i (y - g(\gamma^\top \mathbf{X}_i))$. The main

differences between $\hat{\beta}$ and this AIPW estimator is that we impute $Y_{\tau i}$ by a more flexible model instead of $\hat{Y}_{\tau i}$, and we also exclude the calibration term $\hat{w}_i(Y_{\tau i} - \hat{Y}_{\tau i})$. These differences enable us to further improve efficiency and incorporate regularization easily.

Similar to other IPW methods, the consistency of our proposed τ -GLM estimators relies on the assumption that either C is independent of covariates or $C|\mathbf{X}$ can be modeled correctly via semi-parametric approaches. When both assumptions fail, our proposed estimators are less sensitive to the violation of such assumptions than the simple IPW estimator, as suggested by our numerical results. It is also important to note that additional distributional assumptions about C are required even if model (1) holds for a given τ , due to the curse of dimensionality (Robins and Ritov, 1997). This differs from the standard Cox model, which requires much stronger model assumptions. When both the Cox model and the τ -GLM are misspecified, the proposed estimators have the advantage of being free of censoring, potentially leading to risk models with higher prediction performance.

In our numerical studies, we include spline functions of individual variables for an enriched imputation model. However, other basis functions can be considered provided that X_j is a linear combination of $\Phi(\mathbf{Z})$. In most of our simulations and examples, we only used 3 knots for each variable so that stable estimates could be obtained for the enriched model. A larger number of basis functions can be used for settings where a larger effective sample size is available, although an overly complex model may lead to substantial overfitting and consequently compromise the efficiency gain of the final augmented estimators in the finite sample. The efficiency gain from incorporating auxiliary covariates would depend on how much additional information they have on the outcome, given targeted baseline covariates. For example, if an intermediate variable is highly correlated with some of the baseline covariates, incorporating it using our method would not lead to substantial efficiency gains. It is thus important to select proper surrogate variables in the rare event clinical trial setting.

Post-baseline covariates cannot be used directly in the working baseline prediction model, but we incorporate them in the estimating process to improve the efficiency. However, incorporating post-baseline covariates needs to be handled with caution. In our example using ACTG 175 data, we incorporated covariates collected at week 24 as intermediate variables when less than 0.5% subjects had events and a negligible number of subjects were censored. In practice, this method could be useful in a setting where the post-baseline auxiliary variable is collected at the very early stage of a study, prior to occurrence of events and censoring. One example could be the collection of short-term treatment response markers. However, if collected at a later stage of a study, an intermediate variable would be missing for a non-negligible portion of subjects who already experienced events or dropped out of the study prior to the measurement time. The proposed methods are not directly applicable to such settings and further research is warranted to incorporate short-term outcomes that are subject to censoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank AIDS Clinical Trials Group for providing the data for use in this example. We are also grateful to editors and referees whose comments help to improve the article. This research was partially supported by NIH grants T32AI007358, R01 HL089778, and U54 HG007963.

References

- Bang H, Tsiatis A. Estimating medical cost with censored data. *Biometrika*. 2000; 87:329–343.
- Cai T, Cheng S. Robust combination of multiple diagnostic tests for classifying censored event times. *Biostatistics*. 2008; 9:216–233. [PubMed: 18056687]
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972:187–220.
- DiRienzo G. Flexible regression model selection for survival probabilities: with application to aids. *Biometrics*. 2009; 65:1194–1202. [PubMed: 19173693]
- Friedman, J., Hastie, T., Tibshirani, R. The elements of statistical learning. Vol. 1. Springer; Berlin: 2001. Springer series in statistics
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. 2010; 33:1. [PubMed: 20808728]
- Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical journal*. 2008; 50:457–479. [PubMed: 18663757]
- Gray R. A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika*. 1994; 81:527–539.
- Hammer S, Katzenstein D, Hughes M, Gundacher H, Schooley R, R H, et al. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*. 1996; 335:1081–1090. [PubMed: 8813038]
- Lu X, Tsiatis A. Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrika*. 2008; 95:679–694.
- Matsouaka RA, Li J, Cai T. Evaluating marker-guided treatment selection strategies. *Biometrics*. 2014; 70:489–499. [PubMed: 24779731]
- Parast L, Tian L, Cai T. Landmark estimation of survival and treatment effect in a randomized clinical trial. *Journal of American Statistical Association*. 2014; 109(505):384–394.
- Robins J, A R, Zhao L. Estimation of regression coefficients when some of the regressors are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
- Robins J, Ritov Y. A curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models. *Statistics in Medicine*. 1997; 16:285–319. [PubMed: 9004398]
- Robins J, Seud M, Q L-G, Rotnitzky A. Comment: performance of double-robust estimators when 'inverse probability' weights are highly variable. *Statistical Science*. 2007; 22(4):544–559.
- Robins J, Wang N. Inference for imputation estimators. *Biometrika*. 2000; 87:113–124.
- Robins, JM., Rotnitzky, A. *AIDS Epidemiology*. Springer; 1992. Recovery of information and adjustment for dependent censoring using surrogate markers; p. 297–331.
- Scharfstein D, Rotnitzky A, Robins J. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*. 1999; 94(448):1096–1120.
- Tsiatis, A. *Semiparametric Theory and Missing data*. Springer; New York: 2006.
- Uno H, Cai T, Tian L, Wei L. Evaluating predictionrules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102:527–537.
- Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*. 2007; 34:70–85.
- Zhang HH, Lu W. Adaptive lasso for Cox's proportional hazards model. *Biometrika*. 2007; 94:691–703.
- Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429.

Figure 1.
Simulation results for regularized baseline models

Table 1

Empirical bias, SE (ESE), average of the estimated SE (ASE) and coverage probabilities (CovP) of the 95% CIs for the low-dimensional setting. Shown also are the percent of efficiency gain (%EffG) relative to the IPW_{KM} estimator.

σ_s	Bias $\times 100$					ESE _{ASE} $\times 100$					CovP					%EffG				
	β_0	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4	β_5	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
Low Event Rate																				
0.5	IPW _{KM}	0.8	-0.7	-1.1	-1.4	15.6 _{16.2}	20.0 _{20.0}	20.5 _{19.6}	19.8 _{20.1}	97	96	95	96	96	96	-	-	-	-	-
	AUG _X	1.7	-1.3	-2.0	-0.8	16.0 _{16.0}	17.2 _{17.2}	16.2 _{15.6}	16.4 _{17.3}	96	96	95	97	97	97	-5.9	35.4	57.7	47.0	47.0
	IPW _{cox}	-0.9	-1.0	-1.1	-0.9	14.9 _{17.4}	18.6 _{18.4}	18.2 _{17.5}	18.3 _{18.6}	98	95	96	95	96	95	10.3	16.2	26.3	18.3	18.3
	AIPW _{KM}	1.9	0.8	-1.0	-0.5	15.9 _{16.9}	19.3 _{19.5}	20.9 _{20.3}	19.3 _{19.7}	96	95	93	95	93	95	-4.0	7.4	-4.1	6.4	6.4
	AIPW _{cox}	0.8	0.6	-1.0	-0.5	15.8 _{16.8}	18.6 _{19.0}	19.7 _{19.4}	18.6 _{19.3}	96	94	94	95	94	95	-2.3	15.4	8.1	14.1	14.1
	AUG _{X,S}	2.3	-0.7	-2.8	-0.9	13.5 _{13.2}	14.2 _{13.6}	14.5 _{13.0}	13.9 _{13.6}	95	93	93	95	93	95	30.3	98.0	92.0	103.5	103.5
	AUG _{CMB}	1.8	-0.2	-2.8	-0.7	13.5 _{13.2}	14.0 _{13.4}	14.4 _{12.9}	13.9 _{13.6}	95	93	93	94	93	94	31.3	103.7	96.7	105.8	105.8
	IPW _{cox}	-1.1	-0.9	-1.1	-1.0	15.2 _{17.7}	18.7 _{18.4}	17.9 _{17.5}	18.4 _{18.6}	98	95	96	95	96	95	5.3	13.1	26.7	16.2	16.2
	AIPW _{KM}	1.3	0.4	-1.5	-0.9	15.9 _{17.1}	19.7 _{20.6}	20.4 _{20.8}	19.9 _{20.9}	97	96	95	97	96	95	-4.1	2.8	-2.4	-1.1	-1.1
	AIPW _{cox}	-0.3	-0.0	-1.5	-0.7	15.8 _{16.8}	18.8 _{19.8}	18.8 _{19.4}	19.0 _{20.1}	96	96	96	96	96	96	-2.3	12.1	14.0	9.2	9.2
2	AUG _{X,S}	2.4	-0.7	-2.7	-0.6	15.5 _{15.3}	16.2 _{16.0}	15.6 _{14.8}	15.9 _{16.1}	94	95	93	96	93	96	-0.3	50.5	63.2	54.8	54.8
	AUG _{CMB}	1.4	-0.3	-2.8	-0.2	15.2 _{15.0}	15.9 _{15.8}	15.4 _{14.7}	15.7 _{16.0}	94	95	94	96	94	96	4.7	56.4	67.8	60.2	60.2
	Moderate Event rate																			
	IPW _{KM}	0.2	0.0	-0.4	-1.0	11.6 _{11.7}	14.1 _{14.0}	13.7 _{13.9}	14.1 _{14.1}	95	96	96	94	95	96	-	-	-	-	-
	AUG _X	0.4	-0.6	-0.8	-0.1	11.3 _{11.5}	12.5 _{12.7}	11.6 _{12.0}	12.6 _{12.7}	95	96	96	95	96	95	4.3	25.6	39.8	26.6	26.6
	IPW _{cox}	-0.4	0.0	-0.2	-0.6	11.3 _{12.2}	13.6 _{13.2}	12.4 _{12.9}	13.3 _{13.3}	97	94	96	94	96	94	4.0	7.0	21.9	13.0	13.0
	AIPW _{KM}	0.5	-0.0	-0.7	-0.6	11.2 _{11.7}	13.3 _{13.6}	14.0 _{14.1}	13.2 _{13.7}	95	95	94	96	95	94	6.2	11.7	-3.5	14.5	14.5
	AIPW _{cox}	-0.1	0.0	-0.5	-0.4	11.1 _{11.6}	13.2 _{13.3}	13.0 _{13.5}	12.8 _{13.4}	95	96	96	96	96	96	8.9	13.9	11.3	21.9	21.9
	AUG _{X,S}	0.5	-0.9	-0.7	0.3	10.1 _{10.3}	11.5 _{11.1}	10.8 _{10.7}	11.4 _{11.1}	96	94	95	94	95	94	30.7	49.4	59.9	54.0	54.0
	AUG _{CMB}	0.3	-0.8	-0.7	0.3	10.1 _{10.3}	11.4 _{11.0}	10.8 _{10.6}	11.4 _{11.0}	96	94	94	94	94	94	31.0	50.7	61.2	54.4	54.4

σ_s	Bias $\times 100$				ESE _{ASE} $\times 100$				CovP				%ERG				
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	
2	IPW _{cox}	-0.3	0.0	-0.2	-0.5	11.4 _{12.3}	13.6 _{13.2}	12.4 _{12.9}	13.3 _{13.3}	97	94	96	94	1.9	6.9	22.1	13.3
	AIPW _{KM}	0.5	-0.1	-0.7	-0.7	11.3 _{11.9}	13.7 _{14.1}	13.9 _{14.2}	13.7 _{14.2}	95	96	95	95	3.4	4.5	-3.6	6.4
	AIPW _{cox}	-0.2	-0.1	-0.5	-0.4	11.2 _{11.7}	13.5 _{13.6}	12.8 _{13.4}	13.1 _{13.7}	96	95	97	96	5.7	9.0	14.1	17.1
	AUG _{X-S}	0.7	-0.7	-1.3	0.0	10.9 _{11.2}	12.1 _{12.2}	11.8 _{11.6}	12.1 _{12.2}	96	95	95	94	10.7	33.6	33.7	36.0
	AUG _{CMB}	0.4	-0.6	-1.3	0.1	10.9 _{11.1}	12.2 _{12.1}	11.8 _{11.6}	12.2 _{12.2}	95	95	95	95	11.3	33.3	33.7	35.4

Table 2

Empirical bias, standard errors (ESE), average of the estimated SE (ASE) and coverage probabilities (CovP) of the 95% confidence intervals for covariate-specific risk, $\pi_a(\mathbf{x})=g(\beta_{\tau,a}^T \mathbf{x})$ within each treatment group $A = a$ and the ITE $D(\mathbf{x})$. Shown also are the percent of efficiency gain (%EffG) relative to the IPW_{KM} estimator. All numbers have been multiplied by 100.

	$\pi_0(\mathbf{x})$					$\pi_1(\mathbf{x})$					$D(\mathbf{x})$				
	Bias	ESE _{ASE}	CovP(%)	%EffG	Bias	ESE _{ASE}	CovP(%)	%EffG	Bias	ESE _{ASE}	CovP(%)	%EffG	Bias	ESE _{ASE}	%EffG
Example 1: $\mathbf{x} = (1, 1, 1, 1)^T$															
IPW _{KM}	0.35	5.37 _{5.36}	94	—	0.31	4.61 _{4.68}	96	—	0.04	6.96 _{7.17}	96	—	—	—	—
IPW _{COX}	-0.02	4.72 _{4.99}	96	30.16	0.02	4.13 _{4.38}	95	25.29	-0.04	6.15 _{6.68}	97	28.36	—	—	—
AIPW _{KM}	0.81	5.90 _{5.68}	94	-18.42	0.51	4.86 _{5.04}	95	-10.29	0.30	7.35 _{7.63}	96	-10.38	—	—	—
AIPW _{COX}	0.49	5.40 _{5.27}	94	-1.65	0.25	4.53 _{4.72}	95	3.72	0.24	6.78 _{7.12}	96	5.24	—	—	—
AUG _X	0.05	4.37 _{4.44}	96	51.30	0.05	4.02 _{4.00}	93	32.12	-0.00	5.76 _{6.01}	96	46.44	—	—	—
AUG _{SS}	0.19	4.20 _{3.98}	93	64.19	0.02	3.81 _{3.60}	93	47.19	0.17	5.51 _{5.38}	95	59.41	—	—	—
AUG _{CMB}	0.15	4.17 _{3.93}	93	66.27	-0.01	3.75 _{3.54}	93	52.17	0.16	5.46 _{5.34}	95	62.86	—	—	—
Example 2: $\mathbf{x} = (1, -0.5, -0.5, -0.5)^T$															
IPW _{KM}	0.52	4.57 _{4.40}	94	—	0.45	3.96 _{3.98}	95	0.00	0.07	6.00 _{5.95}	95	0.00	—	—	—
IPW _{COX}	0.15	4.18 _{4.46}	95	20.83	0.23	3.82 _{4.02}	96	8.30	-0.08	5.68 _{6.02}	97	11.60	—	—	—
AIPW _{KM}	0.46	4.56 _{4.53}	95	0.78	0.56	4.04 _{4.14}	94	-4.58	-0.10	6.16 _{6.15}	95	-5.07	—	—	—
AIPW _{COX}	0.19	4.29 _{4.32}	95	14.57	0.34	4.04 _{4.00}	93	-3.55	-0.16	5.98 _{5.90}	95	0.62	—	—	—
AUG _X	0.76	4.06 _{3.96}	94	23.96	0.71	3.78 _{3.67}	93	7.32	0.06	5.64 _{5.41}	94	13.41	—	—	—
AUG _{SS}	0.81	3.77 _{3.50}	93	42.09	0.90	3.56 _{3.29}	92	17.71	-0.09	5.32 _{4.82}	92	27.45	—	—	—
AUG _{CMB}	0.73	3.72 _{3.47}	93	46.78	0.83	3.53 _{3.25}	92	20.48	-0.10	5.26 _{4.78}	93	30.04	—	—	—

Estimated intercept (Int) and covariate effects for the week 144 progression risk prediction models among those treated with ZVD and with ZVD+DDI arms from ACTG175.

Table 3

Coefficient _s β										% Eff _g				
Int	IV	Age	pZDV	SS	CD4	Int	IV	Age	pZDV	SS	CD4	SS	pZDV	CD4
ZDV arm: Event rate=15%														
IPW _{KM}	-0.31 _{0.79}	0.06 _{0.37}	0.33 _{0.15}	0.14 _{0.27}	-0.23 _{0.34}	-0.75 _{0.16}								
IPW _{COX}	-0.59 _{0.77}	-0.12 _{0.37}	0.38 _{0.15}	0.32 _{0.26}	-0.30 _{0.34}	-0.75 _{0.16}	3.1	0.2	2.9	3.5	1.0	3.5		
AIPW _{KM}	-0.28 _{0.81}	0.07 _{0.41}	0.29 _{0.15}	0.22 _{0.28}	-0.20 _{0.36}	-0.72 _{0.16}	-6.4	-16.8	-4.6	-10.5	-10.7	9.1		
AIPW _{COX}	-0.63 _{0.81}	-0.11 _{0.42}	0.35 _{0.15}	0.40 _{0.28}	-0.27 _{0.36}	-0.71 _{0.16}	-6.2	-19.9	-0.3	-8.7	-11.3	6.5		
AUG _X	-0.78 _{0.63}	0.06 _{0.27}	0.28 _{0.12}	0.10 _{0.19}	-0.11 _{0.24}	-0.55 _{0.13}	54.9	84.7	46.4	103.1	94.4	48.0		
AUG _{X,S}	-0.28 _{0.66}	0.06 _{0.28}	0.27 _{0.13}	0.18 _{0.20}	-0.15 _{0.25}	-0.69 _{0.14}	43.2	74.9	40.7	86.9	86.9	39.3		
AUG _{CMB}	-0.27 _{0.67}	0.06 _{0.27}	0.23 _{0.13}	0.20 _{0.19}	-0.11 _{0.24}	-0.69 _{0.14}	39.3	84.5	40.8	101.1	104.6	32.0		
ZDV+DDI arm: Event rate=12%														
IPW _{KM}	-0.69 _{0.89}	0.31 _{0.42}	0.21 _{0.19}	0.49 _{0.33}	1.22 _{0.34}	-0.85 _{0.20}								
IPW _{COX}	-0.96 _{0.90}	0.16 _{0.42}	0.22 _{0.19}	0.71 _{0.33}	1.27 _{0.33}	-0.81 _{0.20}	-0.6	-0.3	3.5	1.6	2.8	-1.8		
AIPW _{KM}	-0.56 _{0.96}	0.33 _{0.48}	0.20 _{0.19}	0.50 _{0.37}	1.19 _{0.34}	-0.86 _{0.21}	-12.3	-20.7	0.7	-16.5	-1.1	-15.1		
AIPW _{COX}	-0.87 _{0.96}	0.19 _{0.48}	0.21 _{0.19}	0.72 _{0.37}	1.27 _{0.34}	-0.82 _{0.21}	-13.5	-20.9	1.5	-19.3	-0.1	-16.9		
AUG _X	-0.62 _{0.78}	0.33 _{0.35}	0.18 _{0.16}	0.50 _{0.29}	1.19 _{0.33}	-0.83 _{0.21}	33.0	45.5	43.1	33.3	3.0	-14.8		
AUG _{X,S}	-0.94 _{0.70}	0.21 _{0.31}	0.19 _{0.15}	0.34 _{0.24}	0.88 _{0.32}	-0.64 _{0.18}	64.9	85.9	70.3	91.1	13.6	24.3		
AUG _{CMB}	-1.04 _{0.70}	0.15 _{0.30}	0.17 _{0.14}	0.30 _{0.24}	0.97 _{0.32}	-0.69 _{0.18}	61.9	94.8	79.9	88.0	11.2	20.1		

Table 4

Predicted progression risk at week 144 among those treated with ZDV/DDI ($A = 1$) and ZVD ($A = 0$), denoted by $\{\pi_a(\mathbf{x}), a = 0, 1\}$, for two examples of covariates and along with their treatment effects using ACTG 175 study.

	ZDV+DDI			ZDV			Treatment Difference		
	$\pi_1(\mathbf{x})$	SE	%EFG	$\pi_0(\mathbf{x})$	SE	%EFG	$D(\mathbf{x})$	SE	%EFG
Example 1: IV=No, Age=22, pZDV=No, SS=No, CD4=100									
IPW _{KM}	0.26	0.0953	–	0.42	0.1113	–	–0.16	0.15	–
IPW _{COX}	0.22	0.0870	20.00	0.38	0.1071	7.89	–0.16	0.1383	12.95
AIPW _{KM}	0.27	0.1058	–18.91	0.41	0.1112	0.15	–0.14	0.1520	–6.58
AIPW _{COX}	0.23	0.0955	–0.44	0.36	0.1072	7.66	–0.14	0.1425	6.33
Cox	0.14	0.0617	138.38	0.39	0.1094	3.43	–0.24	0.1247	38.79
AUG _X	0.26	0.0789	45.79	0.33	0.0858	68.07	–0.07	0.1168	58.13
AUG _{X-S}	0.24	0.0727	71.69	0.40	0.0929	43.57	–0.17	0.1179	55.33
AUG _{CMB}	0.21	0.0715	77.53	0.39	0.0924	45.01	–0.18	0.1165	59.19
Example 2: IV=Yes, Age=40, pZDV=Yes, SS=No, CD4=350									
IPW _{KM}	0.11	0.0449	–	0.19	0.0531	–	–0.07	0.0692	–
IPW _{COX}	0.11	0.0449	–0.11	0.19	0.0515	6.42	–0.07	0.0683	2.67
AIPW _{KM}	0.12	0.0542	–31.26	0.20	0.0622	–27.02	–0.08	0.0805	–26.19
AIPW _{COX}	0.12	0.0514	–23.58	0.19	0.0612	–24.78	–0.07	0.0779	–21.10
Cox	0.12	0.0465	–6.57	0.19	0.0581	–16.39	–0.06	0.0752	–15.39
AUG _X	0.12	0.0399	26.82	0.20	0.0417	62.46	–0.07	0.0569	47.68
AUG _{X-S}	0.13	0.0375	43.12	0.20	0.0431	51.95	–0.07	0.0561	51.87
AUG _{CMB}	0.09	0.0368	48.60	0.18	0.0412	65.83	–0.09	0.0540	64.27