Published in final edited form as:

Biometrics. 2016 September ; 72(3): 742–750. doi:10.1111/biom.12472.

Estimating Treatment Effect in a Proportional Hazards Model in Randomized Clinical Trials with All-or-Nothing Compliance

Shuli Li and

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, U.S.A

Robert J. Gray

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, U.S.A

Shuli Li: shuli@jimmy.harvard.edu; Robert J. Gray: gray@jimmy.harvard.edu

Summary

We consider methods for estimating the treatment effect and/or the covariate by treatment interaction effect in a randomized clinical trial under non-compliance with time-to-event outcome. As in Cuzick et al. (2007), assuming that the patient population consists of three (possibly latent) subgroups based on treatment preference: the *ambivalent* group, the *insisters* and the *refusers*, we estimate the effects among the *ambivalent* group. The parameters have causal interpretations under standard assumptions. The paper contains two main contributions. First, we propose a weighted per-protocol (Wtd PP) estimator through incorporating time-varying weights in a proportional hazards model. In the second part of the paper, under the model considered in Cuzick et al. (2007), we propose an EM algorithm to maximize a full likelihood (FL) as well as the pseudo likelihood (PL) considered in Cuzick et al. (2007). The E step of the algorithm involves computing the conditional expectation of a linear function of the latent membership, and the main advantage of the EM algorithm is that the risk parameters can be updated by fitting a weighted Cox model using standard software and the baseline hazard can be updated using closed form solutions. Simulations show that the EM algorithm is computationally much more efficient than directly maximizing the observed likelihood. The main advantage of the Wtd PP approach is that it is more robust to model misspecifications among the *insisters* and *refusers* since the outcome model does not impose distributional assumptions among these two groups.

Keywords

All-or-Nothing Compliance; EM Algorithm; Proportional Hazards Model; Randomized Clinical Trial; Weighted Partial Likelihood

Correspondence to: Shuli Li, shuli@jimmy.harvard.edu; Robert J. Gray, gray@jimmy.harvard.edu.

Supplementary Materials

Web Appendices A–G, referenced in Sections 2 and 3, and R code are available with this paper at the Biometrics website on the Wiley Online Library.

1. Introduction

In a randomized clinical trial with time-to-event outcome, the treatment effect is often estimated using the proportional hazards (PH) model following the intention-to-treat (ITT) principle. When patient non-compliance occurs, the ITT estimator measures the effect of the treatment assignment, rather than the true treatment efficacy. Simple alternatives to the ITT analysis such as the as-treated analysis (AT) and the per-protocol analysis (PP) could also be biased due to potential confounding by measured or unmeasured factors.

Often, more sophisticated models that incorporate covariates and/or treatment by covariate interactions are of interest. Our motivating study for this paper is the PACCT-1 trial, a phase III breast cancer trial led by the ECOG-ACRIN Cancer Research Group. This trial evaluates the use of the OncoType DX Recurrence Score (RS, developed by Genomic Health, Inc. based on expression profiles from 21 genes, Paik et al. (2004)) as a marker to guide treatment decisions. In the study design, low risk patients ($RS < 11$) were treated with standard hormonal therapy (H) and high risk patients ($RS > 25$) were treated with hormonal therapy + chemo therapy (H+C), while the intermediate risk patients ($RS 11-25$) were randomized between H and H+C. To address the primary question whether adding chemotherapy is beneficial for the intermediate risk group, the study was designed to have high power to detect a difference in 5-year invasive disease-free survival of 90% (H+C) vs. 87% (H). Accrual to the study was closed on October 2010 with 6,907 intermediate risk patients randomized. Further classifying the intermediate risk patients into three RS categories with increasing risk: (11 – 15), (16 – 20), (21 – 25), treatment non-compliance was observed in both arms, within each risk category. As risk category increases, the non-compliance rate increases in arm H (5% vs. 6% vs. 12%) but decreases in arm H+C (25% vs. 15% vs. 11%), implying increased preference for C as risk rises. The term non-compliance used here refers to the situation where a patient was randomized to H but was actually treated with H+C, or vice versa.

While the study was designed to use an ITT comparison as the primary analysis, given that relatively high non-compliance may reduce the power of the analysis and may bias the treatment effect towards no difference, further analyses to adjust for non-compliance may be necessary. As retrospective experience suggested that high RS patients may receive greater benefit from chemotherapy (Paik et al. (2006)), it is also of interest to explore the treatment by RS interaction effect.

Approaches to adjust for non-compliance in randomized trials have been proposed in various outcome settings, including time-to-event endpoints. Most methods are formulated to estimate the causal effect of treatment actually received in structural models for counterfactual outcomes. Robins and Tsiatis (1991) estimated the causal treatment effect in a structural accelerated failure time model. However, in cancer clinical trials, treatment effect on the PH scale is more commonly of interest. Loeys and Goetghebeur (2003) proposed estimating the causal PH effect among the treatable subpopulation defined on the potential treatment compliance variable. Cuzick et al. (2007) proposed several estimators to estimate the treatment effect on a PH scale among the *ambivalent* group, a latent subgroup of subjects who do not have a preference to treatment, but only the mixture likelihood method

is general enough to allow covariate effects and allow correlation between the covariates and the treatment preference groups.

The marginal structural Cox model introduced by Robins (2000) and illustrated by Hernán et al. (2001), can be applied in the non-compliance setting to estimate the population average causal effect of treatment actually received using the inverse probability of treatment weighted (IPTW) method. This method does not exploit randomization but rather relies on the no unmeasured confounder assumption (NUCA), which requires that all covariates predicting both treatment and outcome are collected and incorporated in the weights. It was originally developed for complex longitudinal observational studies with time-varying treatment and therefore could incorporate time dependent non-compliance. Both the IPTW method and the mixture likelihood method could incorporate covariates and possible treatment by covariate interactions.

In this article, we propose a weighted per-protocol (Wtd PP) approach to adjust for treatment non-compliance in a PH model. All-or-nothing compliance is assumed. In the PACCT-1 trial, compliance is with respect to the chemotherapy, which is of short duration relative to time to events. Also, almost all patients who start chemotherapy received several cycles (planned number is 4–6). As in Cuzick et al. (2007), we assume that the patient population consists of three subgroups based on patient treatment preference, the *ambivalent* group, the *insisters* and the *refusers*, and our objective is to estimate the treatment effect among the *ambivalent* group, which was also referred to as the *compliers* in the principal stratification context (e.g. Frangakis and Rubin (2002)). Various papers have considered estimating treatment effect among the *ambivalent* group (e.g. Rubin et al. (1996), Abadie (2003)). Since the outcome model specifies effects within the *ambivalent* group without imposing distributional assumptions on the outcomes for the other two groups, this method is potentially more robust to model misspecifications among the *insisters* and *refusers*, although the model used for estimating the weights does restrict the distributions. On the other hand, the likelihood method considered in Cuzick et al. (2007) does restrict the outcome distributions for the *insisters* and *refusers* through the main outcome models.

In the second part of the paper, under the proportional hazards model considered in Cuzick et al. (2007), we propose an EM algorithm to estimate the parameters in a full likelihood (FL), and we also modify the EM algorithm to maximize the pseudo likelihood (PL) considered in Cuzick et al. (2007). The EM algorithm is computationally much more efficient than directly optimizing the likelihood using standard general purpose algorithms, since the risk parameters can be updated using standard software and the baseline hazard parameters can be updated using closed form formulas.

The paper is organized as follows. In Section 2, we introduce the notation and the model. In Section 3, the Wtd PP estimator is developed, and we propose the FL estimator and the PL estimator obtained from the EM algorithms. The performance of the three estimators is evaluated through simulations in Section 4. In Section 5, the methods are illustrated by analyzing an ECOG-ACRIN breast cancer clinical trial. Section 6 gives relevant discussion.

2. Notation and the model

Consider a clinical trial with non-compliance where the interest is to compare an experimental treatment E vs. a standard treatment S. Denote treatment E as 1 and treatment S as 0. Let $\mathbf{Z}_i$ be a vector of baseline covariates, R_i be the randomized arms and assume $P(R_i = 1|\mathbf{z}_i) = \rho$ so $P(R_i = 0|\mathbf{z}_i) = 1 - \rho$, where ρ is known by design and is independent of $\mathbf{Z}_i$. Let A_i be the treatment actually received. Under all-or-nothing compliance both R_i and A_i are binary variables. Let T_i be time to failure, C_i be censoring time, $U_i = \min(T_i, C_i)$ and $\delta_i = \mathbb{I}(T_i \leq C_i)$. We observe n independent and identically distributed (iid) copies $O_i = (U_i, \delta_i, \mathbf{Z}_i, R_i, A_i)$.

As in Cuzick et al. (2007), assume patient population consists of three latent groups determined by patient preference for E: *ambivalent* (*amb*) – subjects who receive whichever treatment they are offered; *insisters* (*ins*) – subjects who always receive E regardless treatment assigned; and *refusers* (*ref*) – subjects who always receive S regardless treatment assigned. Denote this membership as M_i and throughout the paper, we assume $T_i \perp C_i | \mathbf{Z}_i, A_i, M_i$ with $\perp$ denoting independence. That is, C_i is non-informative, conditional on the covariates, the preference groups and the actually received treatment. This treatment preference model assumes no defiers and it is consistent with the principal stratification framework (Frangakis and Rubin (2002)). Note that while treatment effect among *insisters* and *refusers* are not identified as they always only receive one treatment, patients in the *amb* group always receive their randomized treatments therefore they should provide unbiased information regarding the treatment effect. Further discussions regarding the identifiability of treatment effect under non-compliance can be found in Robins and Rotnitzky (2004). We consider the following proportional hazards model:

$$\lambda_T(t|\mathbf{z}, a, M=\text{amb}) = \lambda_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{\eta}(\mathbf{z}, a)), \quad (1)$$

where $\lambda_0(t)$ is the baseline hazard function, $\boldsymbol{\beta} \in \mathbb{R}^P$, and $\boldsymbol{\eta}(\mathbf{z}, a)$ is a function of $\mathbf{z}$ and a . In the PACCT-1 trial, for example, let $\mathbf{Z} = \text{RS}$, $\boldsymbol{\beta}' \boldsymbol{\eta}(\mathbf{z}, a) = \beta_z z + \beta_a a + \beta_{za}(z \times a)$, where $z \times a$ is the interaction between treatment and RS.

Note that although (1) is defined among $M = \text{amb}$, M is not always observable. Instead, by combining R_i and A_i the following four compliance groups are observed: SE – patients randomized to S but receive E (*ins*); ES – patients randomized to E but receive S (*ref*); SS – patients randomized to S and receive S (mixture of *amb* and *ref*); and EE – patients randomized to E and receive E (mixture of *amb* and *ins*). Denote this observed membership as $M_{\text{obs},i}$. It is clear that M_i can only be directly identified for SE and ES groups, but not for SS and EE groups. By directly applying the arguments given in Chapter 23, Fitzmaurice (2009), we show in Web Appendix A that the parameters in (1) have a causal hazard ratio interpretation.

3. Estimation

3.1 A weighted Per-Protocol (Wtd PP) approach

The per-protocol group includes patients with $R = A$. In Web Appendix B, we show that

$$\lambda(t|\mathbf{z}, A_i=a, M_i=\text{amb}) = \lambda(t|\mathbf{z}, R_i=A_i=a) \frac{P(M_i=\text{amb}|T_i=t, \mathbf{z}, R_i=A_i=a)}{P(M_i=\text{amb}|T_i \geq t, \mathbf{z}, R_i=A_i=a)} = \lambda(t|\mathbf{z}, R_i=A_i=a) W_{ia}(t, \mathbf{z}). \quad (2)$$

That is, given $A = a$ and $\mathbf{Z} = \mathbf{z}$, the hazard in the *amb* group can be identified as the hazard in the per-protocol group, weighted by a time-varying factor. In counting process notation, define the failure process $N_i(t) = 1(U_i \leq t, \delta_i = 1)$ and the at-risk process $Y_i(t) = 1(U_i \geq t)$. It follows (1) that in the subset $\{R_i = A_i = a\}$, the process,

$$\begin{aligned} 1(U_i \leq t, \delta_i = 1) - \int_0^t 1(U_i \geq u) \lambda(u|\mathbf{z}, R_i=A_i=a) du \\ = 1(U_i \leq t, \delta_i = 1) - \int_0^t 1(U_i \geq u) \lambda(u|\mathbf{z}, M_i=\text{amb}, R_i=a) / W_{ia}(u, \mathbf{z}) du, \end{aligned}$$

is a martingale. Linking this together with the standard theory for partial likelihood, we therefore propose to estimate β in (1) by fitting a weighted Cox model for the per-protocol

group, with the contribution of a subject in the risk set weighted by $\frac{1}{W_{ia}(t, \mathbf{z})}$. That is, $\hat{\beta}$ is obtained by maximizing the following weighted pseudo-partial likelihood:

$$L_w(\beta) = \prod_{i=1}^n \left\{ \frac{e^{\beta' \eta(\mathbf{Z}_i, A_i)}}{\sum_j \frac{1(R_j=A_j)}{W_{ja}(u, \mathbf{z})} e^{\beta' \eta(\mathbf{Z}_j, A_j)} Y_j(U_i)} \right\}^{\delta_i 1(R_i=A_i)},$$

with corresponding weighted pseudo-score equation

$$\begin{aligned} U_w(\beta) = \sum_{i=1}^n \int_0^\infty 1(R_i=A_i) \{ \eta(\mathbf{Z}_i, A_i) - \\ \frac{\sum_{j=1}^n \frac{1(R_j=A_j)}{W_{ja}(u, \mathbf{z})} \eta(\mathbf{Z}_j, A_j) Y_j(u) e^{\beta' \eta(\mathbf{Z}_j, A_j)}}{\sum_{j=1}^n \frac{1(R_j=A_j)}{W_{ja}(u, \mathbf{z})} Y_j(u) e^{\beta' \eta(\mathbf{Z}_j, A_j)}} \} dN_i(u), \end{aligned}$$

where $\widehat{W_{ja}(u, \mathbf{z})}$ is an estimate of $W_{ja}(t, \mathbf{z})$ defined in (2).

The weights $W_{ia}(t, \mathbf{z})$ are time-varying (depend on T_i) and subject-specific (depend on $\mathbf{Z}_i$). While consistent estimators of $W_{ia}(t, \mathbf{z})$ are needed in order to get consistent estimators for β , obtaining them is not trivial, since the true functional forms of both numerator and

denominator could be complex, especially when $\mathbf{Z}$ is continuous and/or high dimensional. Smoothing techniques could be considered. Alternatively, in Web Appendix C we propose to estimate the numerator and the denominator separately by fitting parametric models for each piece. Note that this approach does not guarantee that the estimators for the weights are consistent (in fact, most likely they will not be). However, through simulations, we show that for reasonable settings, weights obtained through this approach adjust non-compliance fairly well and reasonable estimators of β can be obtained.

Once the $\widehat{W_{ja}(u, \mathbf{z})}$ are obtained, we use the “BFGS” quasi-Newton optimization algorithm to maximize $L_w(\beta)$, and it appears to converge relatively rapidly in our setting.

3.2 The full likelihood (FL) approach

Cuzick et al. (2007) considered the following proportional hazard model:

$$\lambda_T(t|\mathbf{z}, M=m, a) = \lambda_0(t) \exp\{\gamma_m + \beta' \eta(\mathbf{z}, z_0)\}, \quad (3)$$

where $m = ins, ref, amb$, $z_0 = 1(M = amb \text{ and } A = 1)$, and $\gamma_{amb} = 0$. Given the definition of z_0 , this model also only specifies the treatment effect among the *amb* group. Note that (3) assumes that given $\mathbf{Z}$, the hazards of the *ins* and *ref* groups are proportional to the *amb* group. Again, in the PACCT-1 setting, letting $\beta' \eta(\mathbf{z}, z_0) = \beta_z z + \beta_a z_0 + \beta_{za}(z \times z_0)$, where z is the RS score, (3) implies that the hazards in the three groups are:

$$\lambda(t|\mathbf{z}, a) = \begin{cases} \lambda_0(t) \exp(\gamma_{ins} + \beta_z z) & M=ins, \\ \lambda_0(t) \exp(\gamma_{ref} + \beta_z z) & M=ref, \\ \lambda_0(t) \exp\{\beta_z z + \beta_a a + \beta_{za}(z \times a)\} & M=amb. \end{cases}$$

Note that both Z_0 and M are not observable for SS and EE groups. Cuzick et al. (2007) estimated the parameters using a pseudo likelihood (PL) approach. A new PL algorithm is given in the following section. Here, we also consider a FL estimator. As shown in Web Appendix D, the full likelihood still has a tractable form. Other considerations for a FL estimator include: 1) in the FL setting, a likelihood ratio test can be implemented, and 2) the method proposed in Chen and Little (1999) can be applied to estimate the variance of the risk parameters (see Section 3.4 for details).

Given (3), by first writing the conditional joint distribution $f(u_i, \delta_i | \mathbf{z}_i, m, a_i)$ as

$$f(u_i, \delta_i | \mathbf{z}_i, m, a_i) = \lambda_0^{\delta_i}(u_i) \{ \exp(\gamma_m + \beta' \eta(\mathbf{z}_i, z_{0i})) \}^{\delta_i} e^{-\Lambda_0(u_i) \exp(\gamma_m + \beta' \eta(\mathbf{z}_i, z_{0i}))},$$

we show in Web Appendix D that the full likelihood contribution from subject i is:

$$L_i = \begin{cases} [\lambda_0^{\delta_i}(u_i) \tau_{\text{ins},i}^{\delta_i} S_{\text{ins},i}] \times \pi_{\text{ins}}(\mathbf{z}) & i \in SE, \\ [\lambda_0^{\delta_i}(u_i) \tau_{\text{ref},i}^{\delta_i} S_{\text{ref},i}] \times \pi_{\text{ref}}(\mathbf{z}) & i \in ES, \\ \lambda_0^{\delta_i}(u_i) [\pi_{\text{amb}}(\mathbf{z}) \tau_{\text{amb}_{0,i}}^{\delta_i} S_{\text{amb}_{0,i}} + \pi_{\text{ref}}(\mathbf{z}) \tau_{\text{ref},i}^{\delta_i} S_{\text{ref},i}] & i \in SS, \\ \lambda_0^{\delta_i}(u_i) [\pi_{\text{amb}}(\mathbf{z}) \tau_{\text{amb}_{1,i}}^{\delta_i} S_{\text{amb}_{1,i}} + \pi_{\text{ins}}(\mathbf{z}) \tau_{\text{ins},i}^{\delta_i} S_{\text{ins},i}] & i \in EE, \end{cases} \quad (4)$$

where $\lambda_0()$ is the baseline hazard, $\Lambda_0()$ is the baseline cumulative hazard, $\tau_{\text{ins},i} = \exp(\gamma_{\text{ins}} + \beta' \eta(\mathbf{z}, z_0 = 0))$, $\tau_{\text{ref},i} = \exp(\gamma_{\text{ref}} + \beta' \eta(\mathbf{z}, z_0 = 0))$, $\tau_{\text{amb}_{0,i}} = \exp(\beta' \eta(\mathbf{z}, z_0 = 0))$, $\tau_{\text{amb}_{1,i}} = \exp(\beta' \eta(\mathbf{z}, z_0 = 1))$, $S_{m,i} = e^{(-\Lambda_0(u) \tau_{m,i})}$ is the survival function that depends on the corresponding $\tau_{m,i}$ and $\pi_{\text{ins}}(\mathbf{z}) = P(M = \text{ins} | \mathbf{z})$, $\pi_{\text{ref}}(\mathbf{z}) = P(M = \text{ref} | \mathbf{z})$, so $\pi_{\text{amb}}(\mathbf{z}) = P(M = \text{amb} | \mathbf{z}) = 1 - \pi_{\text{ins}}(\mathbf{z}) - \pi_{\text{ref}}(\mathbf{z})$. Following the approach of Breslow (1974), $\Lambda_0()$ is assumed to be a discrete jump function which only jumps at the observed event times, and we denote the set of jump sizes as $\lambda_0\{u_j\}$, $j = 1, \dots, J$. So, writing $\theta = (\gamma_{\text{ins}}, \gamma_{\text{ref}}, \beta)$, the problem is then to maximize $\prod_i L_i$ over $(\theta, \lambda_0\{u_j\}, \pi_m(\mathbf{z}))$, $\theta = (\gamma_m, \beta)$, $m = \text{ins}, \text{ref}$.

The $\pi_m(\mathbf{z})$ can be re-expressed in terms of probabilities of $M_{\text{obs},i}$ that is, $\pi_{\text{ins}}(\mathbf{z})$

$\frac{P(M_{\text{obs},i} = SE | \mathbf{z})}{1 - \rho}$ and $\pi_{\text{ref}}(\mathbf{z}) = \frac{P(M_{\text{obs},i} = ES | \mathbf{z})}{\rho}$. When $\mathbf{Z}$ is low dimensional and categorical, $\pi_m(\mathbf{z})$ are just discrete probabilities. When $\mathbf{Z}$ is continuous and/or high dimensional, for simplicity, parametric models on the observed data $M_{\text{obs},i}$ (say, indexed by finite dimensional parameter α) can be assumed to model $\pi_m(\mathbf{z})$, so the problem is then to maximize $\prod_i L_i$ over the parameter set $(\theta, \lambda_0\{u_j\}, \alpha)$.

3.2.1 The EM algorithm—Let $CO_i = (U_i, \delta_i, \mathbf{Z}_i, M_i, R_i)$ denote the complete data for subject i . Note that since A_i is completely determined by combining M_i and R_i , it is omitted in the complete data vector. In Web Appendix E, we show that the log of the complete data likelihood is the sum of the following two components:

$$\log L_{CO,i}^1 = \begin{cases} \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{ins},i}^{\delta_i} S_{\text{ins},i}\} & i \in SE, \\ \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{ref},i}^{\delta_i} S_{\text{ref},i}\} & i \in ES, \\ 1(M_i = \text{amb}) \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{amb}_{0,i}}^{\delta_i} S_{\text{amb}_{0,i}}\} + 1(M_i = \text{ref}) \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{ref},i}^{\delta_i} S_{\text{ref},i}\} & i \in SS, \\ 1(M_i = \text{amb}) \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{amb}_{1,i}}^{\delta_i} S_{\text{amb}_{1,i}}\} + 1(M_i = \text{ins}) \log\{\lambda_0^{\delta_i}(u_i) \tau_{\text{ins},i}^{\delta_i} S_{\text{ins},i}\} & i \in EE; \end{cases}$$

$$\log L_{CO,i}^2 = \begin{cases} \log\{\pi_{\text{ins}}(\mathbf{z})\} & i \in SE, \\ \log\{\pi_{\text{ref}}(\mathbf{z})\} & i \in ES, \\ 1(M_i = \text{amb}) \log\{\pi_{\text{amb}}(\mathbf{z})\} + 1(M_i = \text{ref}) \log\{\pi_{\text{ref}}(\mathbf{z})\} & i \in SS, \\ 1(M_i = \text{amb}) \log\{\pi_{\text{amb}}(\mathbf{z})\} + 1(M_i = \text{ins}) \log\{\pi_{\text{ins}}(\mathbf{z})\} & i \in EE. \end{cases}$$

Note that $L_{CO,i}^1$ only involves $(\theta, \lambda_0\{u_j\})$ and $L_{CO,i}^2$ only involves $\pi_m(\mathbf{z})$. Detailed procedures of the EM algorithm are given in Web Appendices F (the E step) and G (the M step). Our simulations show that the EM algorithm is computationally very efficient given the existence

of closed form solutions for $\lambda_0\{u_j\}$ (and for $\pi_m(\mathbf{z})$ when appropriate) and being able to update $\boldsymbol{\theta}$ (and possibly $\pi_m(\mathbf{z})$) using standard packages.

3.3 The pseudo likelihood (PL) approach

The FL approach requires the likelihood to be jointly maximized over $(\boldsymbol{\theta}, \lambda_0\{u_j\}, \pi_m(\mathbf{z}))$, where $\pi_m(\mathbf{z})$ is updated during each iteration. Alternatively, as in Cuzick et al. (2007), a pseudo likelihood approach can be considered, by fixing $\pi_m(\mathbf{z})$ at their initial estimates, and the problem is simplified to maximizing the pseudo likelihood over the parameter set $(\boldsymbol{\theta}, \lambda_0\{u_j\})$. The EM algorithm can again be applied. The complete data log likelihood contribution now simplifies to the first component of the complete data log likelihood in the FL setting, $\log L_{CO,i}^1$. In calculating the E step conditional expectation (and the weights), $\pi_m^k(\mathbf{z})$ are replaced by $\hat{\pi}_m(\mathbf{z})$, where $\hat{\pi}_m(\mathbf{z})$ are fixed over all iterations. The same M step procedure from the FL setting can then be used to update $(\boldsymbol{\theta}, \lambda_0\{u_j\})$.

3.3.1 Estimating $\pi_m(\mathbf{z})$, $m = \text{ins, ref}$ —As discussed in Section 3.2, by expressing these probabilities as probabilities of the observed data $M_{obs,i}$ multinomial/logistic models can be used to obtain their estimates, or when $\mathbf{Z}$ is categorical, the formulas given for choosing the initial values for these probabilities in the FL setting can be used to obtain the estimates.

3.4 Variance estimator

The full likelihood (FL) approach fits into the framework of a non-parametric maximum likelihood (NPML) method for censored data, considered in Zeng and Lin (2007) and Murphy and Van der Vaart (2000). Analogous to the parametric likelihood method, variance of the NPML estimators can be estimated by the inverse of the observed Fisher information matrix, which is usually of very high dimension. Alternatively, when interest only lies in a low dimensional parameter set $(\hat{\boldsymbol{\theta}} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ in our case), the profile likelihood method (Murphy and Van der Vaart (2000)) can be considered and the variance of the low dimensional parameter can be estimated by inverting the information matrix of the profile likelihood. One potential problem for applying this idea is that analytical forms of the second derivative of the observed profile likelihood is often not available and complex numerical differentiation method may be needed. To handle this problem, in this paper, we adopt the method proposed by Chen and Little (1999) which uses the EM-aided differentiation to approximate the second derivative of the profile likelihood based on the idea that the first derivative of the profile likelihood can be relatively easily obtained within the EM framework. Specifically, we apply the following steps to estimate the variance of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}})$:

1. Calculate $S(\boldsymbol{\theta}; \lambda_0\{u_j\}, \pi_m(\mathbf{z})) = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_i \{(\log L_{CO,i}^1) + \sum_i (\log L_{CO,i}^2)\}$, the complete data score function.
2. Obtain the NPML estimates of the parameter of interest, $\hat{\boldsymbol{\theta}}$, by applying the proposed EM algorithm.

3. Perturb the j th component of $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_j$ by a small amount $d = \frac{1}{n}$, where n is the sample size. Denote $\hat{\boldsymbol{\theta}}_j^* = \hat{\boldsymbol{\theta}}_j + d$, and let $\hat{\boldsymbol{\theta}}_j^*$ denote the resulting vector where the j th component is replaced by $\hat{\boldsymbol{\theta}}_j^*$ whereas all the other components are fixed at the corresponding NPML estimates.
4. Apply the proposed EM algorithm to obtain $\lambda_0\{u_j\}^{\hat{\boldsymbol{\theta}}_j^*}$ and $\pi_m(\mathbf{z})^{\hat{\boldsymbol{\theta}}_j^*}$.
5. Calculate $E(S^+) = E\{S(\hat{\boldsymbol{\theta}}_j^*; \lambda_0\{u_j\}^{\hat{\boldsymbol{\theta}}_j^*}, \pi_m(\mathbf{z})^{\hat{\boldsymbol{\theta}}_j^*}) | (\hat{\boldsymbol{\theta}}_j^*, \lambda_0\{u_j\}^{\hat{\boldsymbol{\theta}}_j^*}, \pi_m(\mathbf{z})^{\hat{\boldsymbol{\theta}}_j^*})\}$.
6. Now, perturb $\hat{\boldsymbol{\theta}}_j$ on the other side, that is, define $\hat{\boldsymbol{\theta}}_j^* = \hat{\boldsymbol{\theta}}_j - d$ and repeat Steps 3–5. (Two-sided perturbation suggested by Chen and Little (1999)). Denote the resulting Score function as $E(S^-)$.
7. The second derivative of the observed log profile likelihood with respect to the j th component of $\boldsymbol{\theta}$ can be approximated using $\{E(S^+) - E(S^-)\}(2d)^{-1}$. Denote this as $I(\boldsymbol{\theta}_j)$.
8. Steps 3–6 are repeated for $j = 1, \dots, J$.
9. The covariance matrix of $\boldsymbol{\theta}$ can be estimated using the inverse of the resulting matrix, that is, $(I(\boldsymbol{\theta}_1), \dots, I(\boldsymbol{\theta}_J))^{-1}$.

Given that the Wtd PP estimator involves parameters that increase with sample size (γ_t

embedded in $\frac{1}{W_{ia}(t, \mathbf{z})}$), and both the PL estimator and the Wtd PP estimator involve pseudo terms ($\hat{\pi}_m(\mathbf{z})$ and $\widehat{W_{ia}(t, \mathbf{z})}$, respectively), deriving the asymptotic properties of these estimators could be quite involved and has not been attempted. The general inference results given in Zeng and Lin (2007) for semiparametric likelihood models involving infinite dimensional parameters and the profile approach considered in Murphy and Van der Vaart (2000) could be potential future directions for this. Here, we propose to use bootstrap techniques to estimate the variance of the estimators from these two methods. Let $\mathbf{X} = (X_1, \dots, X_n)$ represents the originally observed data, where $X_i = (U_i, \delta_i, \mathbf{Z}_i, R_i, A_i)$. Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\mathbf{X})$ be the estimator of $\boldsymbol{\beta}$ obtained from data $\mathbf{X}$ using a particular approach. Draw B bootstrap resamples of size n , $\mathbf{X}^{*(1)}, \mathbf{X}^{*(2)}, \dots, \mathbf{X}^{*(B)}$, from the original data. Then the variance of $\hat{\boldsymbol{\beta}}$ can be estimated using the bootstrap variance estimator

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \frac{1}{B-1} \sum_{b=1}^B \{\hat{\boldsymbol{\beta}}_b^* - \bar{\hat{\boldsymbol{\beta}}}^*\}^2,$$

where $\hat{\boldsymbol{\beta}}_b^* = \hat{\boldsymbol{\beta}}(\mathbf{X}^{*(b)})$, $b = 1, \dots, B$, is the estimate computed from the b th bootstrap sample, and $\bar{\hat{\boldsymbol{\beta}}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\boldsymbol{\beta}}_b^*$.

4. Simulation results

Simulations are run under the PACCT-1 setting and three sets of conditions are considered. In Simulation set 1, failure times are generated assuming the three preference groups have proportional hazards and three configurations within the *amb* group are considered: 1) there is a treatment by RS interaction; 2) the interaction effect is absent but a treatment effect is present; and 3) both the interaction effect and the treatment effect are null. In Simulation set 2, hazards in the three groups are assumed to be non-proportional: *ins* and *ref* groups follow Weibull distributions while the hazard for the *amb* group follows exponential distribution. Only the RS score by treatment interaction model is considered in this setting. In Simulation set 3, a more complex proportional hazards model is considered allowing interaction effects between the preference groups and the RS score. In all simulations, the risk parameters are chosen such that the hazard ratios are approximately consistent with the assumptions used in the PACCT-1 study design (except for the null model), but a higher baseline hazard is used. Under each condition, 300 datasets are simulated and for each simulated dataset, a sample size of 1000 is used. When appropriate, bootstrap variance estimators are calculated using $B = 500$.

4.1 Simulation set 1

The RS scores, $Z_i = 11, 12, \dots, 25$, are generated based on the empirical distributions observed on PACCT-1; M_i are generated from multinomial distributions using the following three sets of parameters: (*ins* = 5%, *ref* = 25%, *amb* = 70%), (6%, 15%, 79%) and (12%, 11%, 77%), which correspond to the observed proportions in the three risk groups, respectively: RS scores 11–15, 16–20 and 21–25; R_i follow bernoulli (0.5) and A_i are determined from M_i and R_i : $A_i = 0$ if $M_i = \text{ref}$; $A_i = 1$ if $M_i = \text{ins}$; $A_i = 0$ if $M_i = \text{amb}$ and $R_i = 0$; and $A_i = 1$ if $M_i = \text{amb}$ and $R_i = 1$; T_i follows model (3), where $\lambda_T(t|z, m, z_0) = \lambda_0(t) \exp(\gamma_{\text{ins}} 1(M = \text{ins}) + \gamma_{\text{ref}} 1(M = \text{ref}) + \beta_z z + \beta_a z_0 + \beta_{za}(z \times z_0))$. Three sets of true β parameters are considered: $(\beta_z, \beta_a, \beta_{za}) = (0.1, -0.5, 0.03)$, $(0.1, -0.28, 0)$, and $(0.1, 0, 0)$, with γ parameters fixed at $(\gamma_{\text{ins}}, \gamma_{\text{ref}}) = (0.5, -0.5)$. Both T_i and C_i follow exponential distributions with baseline hazards $\lambda_0(t) = 0.021$ and $\lambda_C(c) = 0.0053$, respectively, which gives approximately 14% censoring rate under the non-null models.

The parameters $\pi_m(z)$, $m = \text{ins}, \text{ref}$, are updated using the closed form solutions for the FL estimator, and are estimated using sample proportions within each of the categories of the RS for the PL estimator: $Z_i = 15, 16 \dots Z_i = 20$, and $Z_i = 21$. The time-varying weights $W_{ia}(t, z)$ in (2) are estimated by fitting multinomial models of $M_{\text{obs},i}$. RS is treated as continuous and an additive model of t and z is fit for estimating the denominator; in estimating the numerator, RS is again treated as a three category variable, and at any given observed event time t , sample proportions within each category of Z are used for estimation.

In addition to the Wtd PP estimator and the two likelihood estimators, we also considered the naïve unweighted per-protocol estimator, and the ITT estimator. Also, for the first data configuration where both a treatment effect and an interaction effect are present, we also obtained the Wtd PP estimator using the true weights calculated from the data generating mechanism, to assess the bias and efficiency loss associated with estimating the weights.

The R function `optim()` was used to maximize the objective functions for the Wtd PP approach, and the R function `coxph()` was used to fit the Cox model for the other methods, including the weighted Cox model considered in the likelihood methods. For the Wtd PP approach, initial values for the parameters were obtained by fitting the Cox model $\lambda(t|z, a) = \lambda_0(t) \exp(\beta_{zz}z + \beta_{aa}a + \beta_{za}(z \times a))$ among the observed compliers. For the likelihood methods, initial values θ , $\lambda_0(t)$, and the $\pi_m(z)$ were obtained as described in Web Appendix G.

In all following results tables, the Emean columns give the empirical means $\hat{\beta}$ from the 300 simulations, Evar give the empirical variance, and $\widehat{\text{var}}$ is the average of the variance estimators from the 300 simulations (using the EM-aided differentiation techniques for the FL estimator, and bootstrap techniques for the Wtd PP estimator and the PL estimator). Under all settings, the bootstrap variance estimators appear to estimate the variance well. The variance estimators from EM-aided differentiation seems to slightly under-estimate the variance.

Under all settings, the performance of the two likelihood estimators is very similar, both are nearly unbiased and have similar variances. In most settings, the likelihood estimators appear to be more efficient than the Wtd PP estimator using estimated weights. However, when true weights are used (results under the treatment by RS interaction model, Table 1), the Wtd PP estimator is also unbiased and appears to be the most efficient estimator. When the treatment effect or the interaction effect is non-null (Table 1 and 2), the naïve per-protocol and the ITT estimators under-estimate the effect size. However, when both the treatment effect and interaction are null (Table 3), the ITT estimator is nearly unbiased and is the most efficient method, while the naïve per-protocol estimator gives the most biased estimate for the treatment effect.

4.2 Simulation set 2

Failure times for the *amb* group follow the same exponential distribution as in Simulation set 1, $\lambda(t|z, a, M = amb) = 0.021 \exp(0.1z - 0.5a + 0.03(z \times a))$. Hazards for *insisters* and *refusers* now follow Weibull proportional hazards model given RS score with baseline

hazards $\lambda_0(t|ins) = \left(\frac{2}{25}\right) \left(\frac{t}{25}\right)^{2-1}$ and $\lambda_0(t|ref) = \left(\frac{5}{60}\right) \left(\frac{t}{60}\right)^{5-1}$, respectively. Let HR_1 and HR_2 denote the ratios of the hazards in the *ins* and the *ref* groups, relative to the hazard in the untreated *amb* group, respectively. Given this data configuration, at $t = (20, 40, 60, 80, 100)$, $(HR_1, HR_2) = (5, 0.00002), (10, 0.0003), (15, 0.001), (20, 0.004), (25, 0.01)$, respectively. Under this relatively extreme non-proportional hazards assumption, the Wtd PP approach still gives reasonable parameter estimates whereas both likelihood methods now give much more biased estimates (Table 4). This is expected since the Wtd PP approach does not directly impose distributional assumptions on the outcomes of *insisters* and *refusers* through the main effect model, whereas the likelihood methods require the failure times of all three groups to have proportional hazards.

4.3 Simulation set 3

Failure times are now generated from the following more complex proportional hazard model allowing for interaction effects between the preference groups and RS:

$$\lambda_T(t|z, m, z_0) = \lambda_0(t) \exp\{\gamma_{ins} I(M=ins) + \gamma_{ref} I(M=ref) + \gamma_{ins,z} (I(M=ins) \times z) + \gamma_{ref,z} (I(M=ref) \times z) + \beta_z z + \beta_a z_0 + \beta_{za} (z \times z_0)\},$$

where $(\gamma_{ins}, \gamma_{ref}, \gamma_{ins,z}, \gamma_{ref,z}) = (0.28, -0.71, 0.03, 0.03)$. The γ parameters are chosen such that the median failure times for the three preference groups are approximately the same as in the other two sets of simulations. All the other parameters remain the same as in the interaction model in Simulation set 1.

The estimation procedure for the Wtd PP approach remains the same. However, the parameters $\tau_{m,i}$ in the likelihood contribution now need to be replaced by $\tau_{ins,i} = \exp(\gamma_{ins} + \gamma_{ins,z}z + \beta_z z)$ and $\tau_{ref,i} = \exp(\gamma_{ref} + \gamma_{ref,z}z + \beta_z z)$. Under this less parsimonious model, in terms of bias, the performance of the Wtd PP estimator and the likelihood estimators is similar to their corresponding counterparts obtained in Simulation set 1. But it appears now that the likelihood estimators are losing their advantages of being more efficient than the Wtd PP estimator due to the fact that the number of risk parameters that need to be estimated is increased. That is, the variances of the likelihood estimators and the Wtd PP estimator are now comparable (Table 5).

5. An example

E1180 (Mansour et al. (1998)) is an ECOG-ACRIN breast cancer study where patients are randomized between the combination chemotherapy CMFP (cyclophosphamide/methotrexate/5 fluorouracil/prednisone) and observation. Time from randomization to recurrence is the primary endpoint. Among the 424 analyzable patients, observed non-compliance rate is 11% (22/208) on the CMFP arm, and 16% (34/216) on the observation arm, respectively. Covariates of interest include estrogen receptor (ER) status (positive vs. negative) and primary tumor size (< 3cm vs. ≥ 3cm). ER positive patients with tumors < 3cm were not eligible.

Define two binary variables $Z_1 = 1(\text{ER} = \text{neg}, \text{tumsz} > 3\text{cm})$, and $Z_2 = 1(\text{ER} = \text{pos}, \text{tumsz} > 3\text{cm})$. For the Wtd PP method we assume the following model, $\lambda(t|Z_1, Z_2, a, M_i = \text{amb}) = \lambda_0(t) \exp(\beta_{Z_1} Z_1 + \beta_{Z_2} Z_2 + \beta_a a)$ and for the mixture likelihood methods we assume $\lambda(t|Z_1, Z_2, m, a) = \lambda_0(t) \exp(\beta_{Z_1} Z_1 + \beta_{Z_2} Z_2 + \gamma_m + \beta_a a)$, where $m = \text{ins}, \text{ref}, \text{amb}$, $\gamma_{\text{amb}} = 0$, and $z_0 = 1(M = \text{amb} \text{ and } A = 1)$. The results given in Table 6, show that the treatment effect estimates and their bootstrap standard errors from the Wtd PP approach and the PL method are very close. The performance of the two likelihood estimators is also similar.

6. Discussion

We consider methods for estimating the treatment effects (and potentially the treatment by covariate interaction effects) in randomized clinical trials under all-or-nothing compliance for time-to-event outcomes. We propose a weighed per-protocol (Wtd PP) method which is potentially more robust to certain model misspecifications on the outcomes than a likelihood approach. The EM algorithms we proposed for the likelihood estimators are

computationally more efficient than directly maximizing the observed likelihood over the high dimensional parameter set.

When true weights are known, the Wtd PP estimator gives unbiased estimates that are more efficient than the likelihood estimators. But in reality, the weights need to be estimated. When this is the case, the Wtd PP estimator that uses estimated weights from fitting parametric model appears to be less efficient, and more biased than the likelihood estimators when the likelihood is correctly specified. However, the efficiency differences also appear to depend on the correct model giving a parsimonious relationships among the distributions of the *ambivalent*, *insister*, and *refuser* groups. Obtaining better estimates of the time-varying weights in the Wtd PP estimator is challenging but may worth exploring in future work.

To adjust for non-compliance in a randomized trial setting, the IPTW estimator in marginal structural models that was developed for observational studies could also be considered. But these methods do not exploit randomization and also the no unmeasured confounder assumption (NUCA) is rather strong. Although estimators that do not require NUCA and are robust to model misspecifications are certainly desirable, so far attempts made in the literature toward this direction have not yet been completely successful for time-to-event outcomes. As shown in the simulation results, the Wtd PP method proposed here is potentially more robust in the sense that it does not require any assumptions on outcomes of the *insisters* and *refusers* through the main model, and the optimization of the (pseudo) objective function only involves a low dimensional risk parameter set. Based on our experience, the following may be considered as a guideline, with respect to the use of the three methods considered in this paper. The likelihood methods may be considered when the following assumptions can be reasonably made: 1) the outcome depends on the covariates/membership through a fairly simple form (a parsimonious model), and 2) the proportional hazards assumption reasonably holds among the three compliance groups. Otherwise, the Wtd PP method should be considered. The performance of the two likelihood estimators is similar in our settings. While the EM procedure for the PL estimator is computationally faster, when a likelihood ratio test is desirable, or when the Chen and Little (1999) variance estimator is preferred over a bootstrap variance estimator, the FL estimator could be considered. In practice, when non-compliance is substantial, more than one method may need to be considered together to give a better understanding of the data and more appropriate interpretation of the results obtained from each particular method.

In this paper, we focus on estimating the hazard ratio (HR), which has an causal interpretation in our setting. Hernán (2010) discussed the potential limitation of using a single value HR as the effect measure, when the period-specific HRs vary over time and/or there is selection bias over time, and as suggested in the paper, adjusted survival curves could be considered under these situations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for Shuli Li's research was provided by National Institutes of Health doctoral training grant T32CA009337. The authors thank ECOG-ACRIN for permission to use the data.

References

- Abadie A. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*. 2003; 113(2):231–263.
- Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974;89–99. [PubMed: 4813387]
- Chen HY, Little RJA. Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*. 1999;896–908.
- Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(4):565–588.
- Fitzmaurice, GM. Longitudinal data analysis. Chapman & Hall/CRC: 2009.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58(1):21–29. [PubMed: 11890317]
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*. 2001; 96(454):440–448.
- Hernán, Miguel A. The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*. 2010; 21(1):13.
- Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics*. 2003; 59(1):100–105. [PubMed: 12762446]
- Mansour EG, Gray R, Shatila AH, Tormey DC, Cooper MR, Osborne CK, Falkson G. Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis—an intergroup study. *Journal of clinical oncology*. 1998; 16(11):3486–3492. [PubMed: 9817265]
- Murphy SA, Van der Vaart AW. On profile likelihood. *Journal of the American Statistical Association*. 2000;449–465.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine*. 2004; 351(27):2817–2826. [PubMed: 15591335]
- Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology*. 2006; 24(23):3726–3734. [PubMed: 16720680]
- Robins J. Marginal structural models versus structural nested models as tools for causal inference. *Statistical models in epidemiology, the environment and clinical trials*. 2000; 116:95.
- Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika*. 2004; 91(4):763–783.
- Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in statistics-theory and methods*. 1991; 20(8): 2609–2631.
- Rubin DB, Imbens GW, Angrist JD. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 91(434):444–455.
- Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(4):507–564.

Table 1

Parameter and variance estimates under a model assuming the 3 groups have proportional hazards (number of simulations = 300, sample size in each dataset = 1000); true $\beta = (0.1, -0.5, 0.03)$.

Method	$\hat{\beta}_z$			$\hat{\beta}_a$			$\hat{\beta}_{sa}$		
	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$
<i>wtd PP</i>	0.10	0.00037	0.00036	-0.47	0.040	0.038	0.029	0.00061	0.00059
<i>wtd PP true w</i>	0.10	0.00018	-	-0.51	0.022	-	0.030	0.00037	-
<i>PL</i>	0.10	0.00016	0.00017	-0.51	0.033	0.033	0.031	0.00043	0.00043
<i>FL</i>	0.10	0.00016	0.00015	-0.51	0.034	0.028	0.029	0.00043	0.00040
<i>unwtd PP</i>	0.10	0.00019	-	-0.31	0.024	-	0.025	0.00038	-
<i>ITT</i>	0.11	0.00017	-	-0.34	0.020	-	0.019	0.00032	-

Table 2

Parameter and variance estimates under a model assuming the 3 groups have proportional hazards (number of simulations = 300, sample size in each dataset = 1000); true $\beta = (0.1, -0.28, 0)$

Method	$\hat{\beta}_z$			$\hat{\beta}_a$			$\hat{\beta}_{ca}$		
	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$
<i>wid pp</i>	0.10	0.00030	0.00036	-0.25	0.034	0.037	-0.0020	0.00053	0.00058
<i>PL</i>	0.10	0.00014	0.00016	-0.27	0.031	0.031	-0.0020	0.00043	0.00042
<i>FL</i>	0.10	0.00014	0.00015	-0.28	0.031	0.027	-0.0038	0.00043	0.00039
<i>unwid pp</i>	0.11	0.00016	-	-0.097	0.023	-	-0.0043	0.00038	-
<i>ITT</i>	0.11	0.00016	-	-0.18	0.018	-	-0.0042	0.00031	-

Table 3

Parameter and variance estimates under a model assuming the 3 groups have proportional hazards (number of simulations = 300, sample size in each dataset = 1000); true $\beta = (0.1, 0, 0)$

Method	$\hat{\beta}_z$			$\hat{\beta}_a$		
	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$
<i>wid pp</i>	0.098	0.00013	0.00015	0.011	0.0087	0.0082
<i>PL</i>	0.10	0.000076	0.000095	-0.0050	0.0096	0.0095
<i>FL</i>	0.10	0.000075	0.000091	-0.0087	0.0095	0.0084
<i>unwid pp</i>	0.10	0.000080	-	0.13	0.0057	-
<i>ITT</i>	0.11	0.000070	-	-0.0045	0.0048	-

*: only the additive models of z and a were fit

Parameter and variance estimates under a model assuming the 3 groups have non-proportional hazards (number of simulations = 300, sample size in each dataset = 1000); true $\beta = (0.1, -0.5, 0.03)$

Table 4

$\hat{\beta}_z$			$\hat{\beta}_a$			$\hat{\beta}_{ca}$		
Method	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$	Emean	Evar
<i>wid PP</i>	0.099	0.00033	0.00034	-0.56	0.034	0.035	0.036	0.00057
<i>PL</i>	0.084	0.00012	0.00014	-0.83	0.035	0.040	0.072	0.00046
<i>FL</i>	0.083	0.00011	0.00015	-0.87	0.042	0.030	0.074	0.00051

Parameter and variance estimates under a more complex model allowing for interaction effects between the latent groups and RS score (number of simulations = 300, sample size in each dataset = 1000); true $\beta = (0.1, -0.5, 0.03)$

Table 5

$\hat{\beta}_z$				$\hat{\beta}_a$				$\hat{\beta}_{za}$			
Method	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$	Emean	Evar	$\widehat{var}$	Emean	Evar
<i>wid PP</i>	0.10	0.00038	0.00036	-0.46	0.042	0.040	0.028	0.00063	0.00060		
<i>PL</i>	0.10	0.00036	0.00041	-0.51	0.046	0.047	0.031	0.00064	0.00069		
<i>FL</i>	0.10	0.00036	0.00031	-0.51	0.047	0.036	0.029	0.00065	0.00056		

Table 6

Effect estimates from E1180

Method	$\hat{\beta}_{z_1}(se)$	$\hat{\beta}_{z_2}(se)$	$\hat{\beta}_a(se)$
<i>wtd PP</i>	0.48 (0.34)	0.78 (0.33)	-1.05 (0.33)
<i>PL</i>	0.47 (0.22)	0.66 (0.23)	-1.02 (0.31)
<i>FL</i>	0.49 (0.21)	0.66 (0.22)	-1.03 (0.27)
<i>unwtd PP</i>	0.40 (0.23)	0.68 (0.23)	-0.74 (0.19)
<i>ITT</i>	0.41 (0.20)	0.61 (0.22)	-0.71 (0.18)