HHS Public Access

Author manuscript

J Biopharm Stat. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

J Biopharm Stat. 2016 ; 26(2): 269–279. doi:10.1080/10543406.2014.1000549.

Sensitivity Analysis for Missing Outcomes in Time-to-event Data with Covariate Adjustment

Yue Zhao[*], **Benjamin Saville**[†], **Haibo Zhou**[‡], and **Gary G. Koch**[‡]

[*]Merck Research Laboratories, North Wales, PA

[†]Berry consultants, Austin, TX

[‡]Biostatistics, University of North Carolina, Chapel Hill, NC

Abstract

Covariate-adjusted sensitivity analyses is proposed for missing time-to-event outcomes. The method invokes multiple imputation (MI) for the missing failure times under a variety of specifications regarding the post-withdrawal tendency for having the event of interest. With a clinical trial example, we compared methods of covariance analyses for time-to-event data, i.e., the multivariable Cox proportional hazards model and non-parametric ANCOVA, and then illustrated how to incorporate these methods into the proposed sensitivity analysis for covariate adjustment. The MI methods considered are Kaplan-Meier Multiple Imputation (KMMI), covariate-adjusted and unadjusted proportional hazards multiple imputation (PHMI). The assumptions, statistical issues, and features for these methods are discussed.

Keywords

Time-to-event data; Multiple imputation; Sensitivity analysis; Covariate adjustment

1. Introduction

Missing data exist in practically all clinical trials. A major source of missing data is from patients discontinuing their assigned treatment and then withdrawing from the study. The extent to which missing data impact statistical inferences depends on the process (i.e., mechanism) leading to the missingness. Little and Rubin (2002) outlined the following missing data framework: (i) data are missing completely at random (MCAR) if the missingness does not depend on either the observed or unobserved data; (ii) data are considered missing at random (MAR) when the missingness only depends on the observed data; (iii) data are missing not at random (MNAR) if the missingness depends on the unobserved data. If the measurement process and the missing data process have separate sets of parameters under the MAR mechanism, the missing data mechanism is said to be ignorable for likelihood-based inference since unbiased (or consistent) parameter estimates can be obtained from the observed data (Mallinckrodt et al., 2008).

Corresponding Author: Yue Zhao, M.D., Dr.P.H., BARDS-Late Development Statistics, Merck Research Laboratories, 351 North Sumneytown Pike, North Wales, PA 19454, (267)305-8403, yue.zhao4@merck.com.

In many clinical trials, MAR can be reasonable, and hence it is often chosen as the main assumption for the primary analysis (Mallinckrodt et al. 2008; Zhang, 2009). However, the missing mechanism can be more complex than the ideal MAR assumption in practice. The possibility of MNAR can never be ruled out. Therefore, a prudent analyst should always conduct sensitivity analyses to assess the robustness of the treatment effect inferences to various alternative missing data assumptions (NRC, 2010). Zhao et al. (2014) recently introduced a method for sensitivity analysis for missing outcomes in time-to-event data, for which the primary analytical strategy has the MAR-like assumption of non-informative independent censoring. Based on the Kaplan-Meier (KM) estimator or its Cox proportional hazards (PH) model (Cox, 1972) counterparts, Zhao et al. (2014) employed multiple imputation of potential times to event for withdrawal patients to produce the inference if they were followed off treatment until the end of the study. The departure from the primary MAR-like assumption was addressed by a sensitivity parameter that captures the difference between withdrawal patients and continuing patients for the post-discontinuation tendency of developing an event. When the treatment effects are evaluated with the standard methods without covariate adjustment, application of such a sensitivity analysis is straightforward (Zhao et al., 2014).

Although the unadjusted analysis provides valid treatment comparisons in randomized studies, covariate-adjusted analysis is often implemented to increase statistical power or to offset the influence of random imbalances between treatment groups for the covariates with possibly strong relationships with the primary outcome (Tangen and Koch, 2000). One concern regarding the appropriateness of covariate adjustment with the Cox regression model is whether the proportional hazards assumption holds for each covariate in the model. In addition, incorrect model specifications may produce biased estimates for the regression coefficients (Tangen and Koch, 2000). One way to avoid those issues is to account for the covariates with the randomization based analysis of covariance (ANCOVA). Through weighted least squares methodology (Grizzle et al., 1969), non-parametric approaches have been proposed to provide covariate adjustment for inferences on incidence density ratios (Tangen and Koch, 2000) or hazard ratios (Moodie et al., 2011) for multiple non-overlapping time intervals. Recently, Saville and Koch (2013) discussed a randomization based method to estimate the covariate-adjusted population average hazard ratio with Cox regression models. Using an estimated covariance matrix for the unadjusted log hazard ratio from the Cox regression model and the group differences in means of baseline covariates, they implemented the weighted least squares methodology to produce a covariate-adjusted log hazard ratio by forcing the differences in means for covariables to zero. The central feature of this approach is that it incorporates the usual Cox regression model estimates into the non-parametric ANCOVA (NPANCOVA) paradigm, and hence it avoids the proportional hazards assumption for the adjusted covariates and avoids possibly data driven model refinements. Consequently, it could be an appealing strategy for the primary analysis in regulatory environments.

In this article, we discuss how to implement covariate adjustment in the sensitivity analysis proposed in Zhao et al. (2014) for time-to-event data. When data are from randomized clinical trials, one can regard the patients in each treatment group as comparable to a random sample from the study population. As a result, one straightforward way to perform multiple

imputation (MI) is through Kaplan-Meier (KM) estimates, i.e., the KMMI method in Zhao et al. (2014), mainly because the KM curve is a valid estimator of the survivor profile for a randomized treatment. For each imputed dataset, the covariate-adjusted log hazard ratio can be obtained with the method of Saville and Koch (2013). The final treatment estimate can be obtained from these estimates using Rubin (1987)'s formulas. Alternatively, one can impute data via the Breslow estimator from the Cox proportional hazards model that includes treatment and the set of covariates, and then proceed with analysis by the same Cox regression model.

2. Covariate-Adjusted Hazard Ratio Estimation

Saville and Koch (2013) proposed a non-parametric, randomization-based ANCOVA (NPANCOVA) method to obtain covariate-adjusted log hazard ratios. Let $h = 1,2$ index the test and the control group with n_h patients in group h; and let $\boldsymbol{r}_h$ be the corresponding dfbeta residual ($n_h \times 1$) vector obtained from the unadjusted Cox proportional hazards model with treatments as the only factor. The i-th element of $\boldsymbol{r}_h$ is the change in the log hazard ratio estimate ($\hat{\beta}$) for comparing test treatment versus control when the i-th observation in group h is omitted, and it can be approximated by $-I^{-1}S_{hi}$, where $I(\hat{\beta})$ is the observed information, and S_{hi} is the i-th score residual. Therefore, for $\boldsymbol{r}=(\boldsymbol{r}_1', \boldsymbol{r}_2')'$, $\boldsymbol{r}'\boldsymbol{r}=(\boldsymbol{r}_1'\boldsymbol{r}_1+\boldsymbol{r}_2'\boldsymbol{r}_2)$ approximates the robust sandwich variance for $\hat{\beta}$ (Wei et al., 1989; Lin and Wei, 1989). Let $\boldsymbol{X}_h = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iq})$ be the ($n_h \times q$) matrix for the q baseline covariates for group h; and let $\bar{\boldsymbol{x}}_h=(\boldsymbol{X}_h'\mathbf{1}_{n_h}/n_h)=(\bar{\boldsymbol{x}}_{h1}, \ldots, \bar{\boldsymbol{x}}_{hq})'$ be the vector of means for the q baseline covariates for group h with the corresponding covariance matrix shown in (1),

$$\boldsymbol{V}_{\bar{\boldsymbol{x}}_h}=(\boldsymbol{X}_h-\mathbf{1}_{n_h}\bar{\boldsymbol{x}}_h')'(\boldsymbol{X}_h-\mathbf{1}_{n_h}\bar{\boldsymbol{x}}_h')/(n_h(n_h-1))=\boldsymbol{C}_h'\boldsymbol{C}_h \quad (1)$$

where $\boldsymbol{C}_h=(\boldsymbol{X}_h-\mathbf{1}_{n_h}\bar{\boldsymbol{x}}_h')/\sqrt{n_h(n_h-1)}$ and $\mathbf{1}_{n_h}$ is a ($n_h \times 1$) vector of ones. Let $\boldsymbol{d} = (\hat{\beta}, (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)')'$ be the vector of the unadjusted log hazard ratio estimate for treatments $\hat{\beta}$ and the differences in means for the baseline covariates for the test treatment and the control groups. Then the covariance matrix of $\boldsymbol{d}$ is obtained via the sums of cross products $\boldsymbol{r}_h$ and $\boldsymbol{C}_h$ as shown in (2). The mathematical derivations were discussed by Saville and Koch (2013).

$$\boldsymbol{V}_{\boldsymbol{d}}=\begin{bmatrix} (\boldsymbol{r}_1'\boldsymbol{r}_1+\boldsymbol{r}_2'\boldsymbol{r}_2) & (\boldsymbol{r}_1'\boldsymbol{C}_1-\boldsymbol{r}_2'\boldsymbol{C}_2) \\ (\boldsymbol{r}_1'\boldsymbol{C}_1-\boldsymbol{r}_2'\boldsymbol{C}_2) & (\boldsymbol{C}_1'\boldsymbol{C}_1+\boldsymbol{C}_2'\boldsymbol{C}_2) \end{bmatrix} \quad (2)$$

With the NPANCOVA approach discussed in Koch et al. (1998), the covariate-adjusted estimate for the log hazard ratio can be obtained via the weighted least squares regression (Grizzle et al., 1969) for the model $E_A(\boldsymbol{d}) = \boldsymbol{Z}\delta$, where $E_A(\boldsymbol{d})$ is the asymptotic expected value for $\boldsymbol{d}$, $\boldsymbol{Z}=[1, \mathbf{0}_q']'$ is the matrix to specify the adjusted analysis, and δ is the regression coefficient for the covariate-adjusted log hazard ratio for treatments. With $\boldsymbol{Z}$ to force the difference in means for covariates to zero, the covariate-adjusted log hazard ratio estimate for treatments is $\hat{\delta}=(\boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{d}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{d}}^{-1}\boldsymbol{d}$ and the corresponding variance estimator is

$V_\delta = (\boldsymbol{Z}' \boldsymbol{V}_{\boldsymbol{d}}^{-1} \boldsymbol{Z})^{-1}$. When the sample sizes for each group are sufficiently large for $\boldsymbol{d}$ to have an approximately multivariate normal distribution, confidence intervals and p-values of corresponding statistical tests for the covariate-adjusted log hazard ratio can be based on $(\hat{\delta} - \delta) V_\delta^{-1/2}$ having an approximately normal distribution. The rationale for randomization based covariance adjustment is the expected absence of differences between the test treatment and the control groups for means of the covariables. A related criterion for evaluating the extent of random imbalances between the test treatment and the control groups is Q_0 in (3), which approximately has a chi-square distribution with q degrees of freedom.

$$Q_0 = (\boldsymbol{d} - \boldsymbol{Z}\hat{\delta})' \boldsymbol{V}_{\boldsymbol{d}}^{-1} (\boldsymbol{d} - \boldsymbol{Z}\hat{\delta}) \quad (3)$$

3. Sensitivity Analysis using Multiple Imputation

3.1 Kaplan-Meier Multiple Imputation (KMMI) Strategy

The Kaplan-Meier Multiple Imputation (KMMI) strategy, implemented separately within individual treatment groups, was described in Zhao et al. (2014). Briefly, we assume that a randomized group has events observed at M distinct times ($t_1 < t_2 < \cdots < t_M$), and it has premature discontinuation (i.e., censoring) of patients at K distinct times ($c_1 < c_2 < \cdots < c_K$). With k indexing the censoring times, $t_{k,0}$ denotes the latest failure time prior to c_k and $t_{k,j}$ denotes the j-th failure time after c_k. The imputation scheme is as follows:

i. Obtain the Kaplan-Meier (KM) estimates $\hat{S}(t)$ for the survival distribution with support of $t \in (t_1 < t_2 < \cdots < t_M)$. For the end of follow-up time t^* or the censored times after the latest failure time (i.e., $t_M < c_k < t^*$), an exponential model is used to extrapolate $\hat{S}(t_M)$ to $\hat{S}(t^*)$ or the corresponding $\hat{S}(c_k)$.

ii. With the survivor rate $\hat{S}(c_k)$ (for the patient discontinuing treatment at the time $c_k \leq t_M$) defined by a linear interpolation of $\hat{S}(t_{k,0})$ and $\hat{S}(t_{k,1})$, the probability of having an event in the time interval $[t_{k,j}, t_{k,j+1}]$ conditional on not having the event by the time c_k is given by

$$\hat{f}_{k,j}(\theta) = \frac{\hat{S}(t_{k,j})^\theta - \hat{S}(t_{k,j+1})^\theta}{\hat{S}(c_k)^\theta}, \quad (4)$$

where the *sensitivity parameter* θ is a specified hazard ratio for calibrating the extent to which a patient with premature discontinuation has an event after the censoring time c_k relative to the patients continuing follow-up (while remaining in their randomly assigned treatment group). For alternative sensitivity analyses, specifications for θ can vary between 0 and ∞, with $\theta = 1$ corresponding to the MAR-like assumption of independent random censoring. Specifications of $\theta > 1$ make a prematurely discontinuing patient more likely to have an event after their censoring time than continuing patients, and specifications of $\theta < 1$ have the opposite behavior.

iii. The discontinued patients have their censoring times replaced by the failure times drawn at random from their corresponding conditional distributions with cumulative density function

$$\hat{F}_{k,j}(\theta)=1-\frac{\hat{S}(t_{k,j+1})^{\theta}}{\hat{S}(c_k)^{\theta}}. \quad (5)$$

iv. The imputation procedure is repeated to form L imputed data sets.

The imputed data sets do not have any patient with premature discontinuation, and so analysts can apply the conventional analysis methods for time-to-event data with the censoring only at the end of follow-up time t^*.

A usual way to perform sensitivity analyses with KMMI in each group is to invoke various values of the sensitivity parameter that address different post-discontinuation tendencies of having events. A principle for specifying the sensitivity parameter θ was discussed previously (Zhao et al., 2014). Briefly, analysts could specify θ_T larger than that for the placebo group to penalize the premature discontinuation for the test treatment. With specifying $\theta_P = 1$ for the placebo group to approximate non-informative independent censoring (so that placebo patients with premature discontinuation would have comparable experience after discontinuation to their counterparts without premature discontinuation), $\theta = (\theta_T/\theta_P) = \theta_T$ for the test treatment becomes a single parameter for calibrating sensitivity analyses.

3.2 Covariate adjusted multiple imputation

Covariate-adjusted proportional hazards multiple imputation (PHMI) can proceed for every prematurely discontinued patient by using a patient specific survival distribution estimated by the Breslow estimator from the Cox proportional hazards model with treatments and the set of covariates for the imputation scheme in (ii) – (iv) in Section 3.1 (or the unadjusted model with treatments alone). The covariate-adjusted hazard ratios can then be obtained from imputed data sets by fitting the same Cox regression model for the MI process or by applying NPANCOVA as described in Section 2.

3.3 Parameter estimation

Following well-established rules (Rubin, 1987; Rubin and Schenker, 1991), the method for combining results from L imputed data sets can be applied easily by the SAS procedure MIANALYZE. Let β be the log hazard ratio that would be estimated from the complete data. Let $\hat{\beta}^{(l)}$ denote the point estimate for β and let $\hat{V}_{\beta}^{(l)}$ denote its variance estimate from the l-th data set.

The overall multiple imputation (MI) estimate of β is obtained by averaging the estimates from the L complete-data analysis, $\overline{\beta}=(1/L)\sum_{l=1}^{L}\hat{\beta}^{(l)}$, and its estimated variance is the sum of the within-imputation variance $\overline{V}_{\beta}=(1/L)\sum_{l=1}^{L}\hat{V}_{\beta}^{(l)}$ and the product of the between-

imputation variance $B_\beta=(L-1)^{-1}\sum_{l=1}^{L}(\hat{\beta}^{(l)}-\hat{\beta})^2$ and a finite sample correction shown in (6).

$$\hat{V}_{\overline{\beta}}=\overline{V}_\beta+(1+L^{-1})B_\beta \quad (6)$$

Given sufficiently large sample size for the complete data to support an approximately standard normal $N(0,1)$ distribution for its hypothetical version of $(\hat{\beta}-\beta)\hat{V}_{\hat{\beta}}^{-1/2}$ when there were no missing data, confidence intervals for β (and p-values for corresponding statistical tests) can be based on $(\overline{\beta}-\beta)\hat{V}_{\overline{\beta}}^{-1/2}$ having a t-distribution with approximate degrees of freedom (d.f.) as shown in (7).

$$\begin{aligned}\text{d. f. } &=(L-1)\left(1+\left(\frac{(1+L^{-1})B_\beta}{\overline{V}_\beta}\right)^{-1}\right)^2 \quad (7)\\ &=(L-1)(1+R^{-1})^2\end{aligned}$$

Here, R expresses the relative increase in variance due to missing information.

4. Application

We illustrate the proposed methods with a clinical trial for maintenance treatment for bipolar disorder (Calabrese et al., 2003). For reasons related to confidentiality of the data from this clinical trial, the application uses a data set of 300 patients (150 patients with the test treatment and 150 patients with the placebo) from a random sample (with replacement) from the true study population. The same data set was also used previously in Zhao et al. (2014). After an 8 to 16 weeks run-in period within which all patients received test treatment, eligible patients who tolerated and adhered to the therapy were randomized to the test treatment or to the placebo for a planned follow-up period of 76 weeks. Accordingly, this study had a randomized withdrawal design, and the primary efficacy endpoint was the time-to-intervention for any mood episode. Of the 300 patients, 75 patients (50%) on the test treatment and 82 patients (54.7%) on the placebo had the event of intervention for any mood episode. A total of 97 (32.33%) patients discontinued the study prematurely (35% of the placebo group and 29% on the test treatment).

Seven covariables had *a priori* specification as being of interest in the analysis plan and in the protocol for this clinical trial. Two of them are patients' demographics, and the rest of them are baseline psychiatric assessments related to disease progression in previous studies. The distributions of these covariables are presented in Table 1. The extent of random imbalance between treatments is summarized for each covariate with the standardized difference (i.e., the difference between means divided by the square root of the average of the two sample variances) and the two-sided p-value from the Wilcoxon rank sum test for the association between the covariate and the treatment assignment. The standardized difference (Std. Diff.) represents the difference in means between two groups in units of the standard deviation (STD), and some authors suggest that Std. Diff. < 10% likely expresses a

negligible imbalance (Austin et al., 2010). Table 2 describes the associations between the covariates and the primary endpoint, as assessed by the corresponding Cox regression models stratified on the treatment. Under the assumption of non-informative independent censoring, the univariate analyses for each individual covariate and the multivariate regression analysis were used to evaluate associations for the covariates. Of the five covariates with Std. Diff. ≥ 10%, the pre-randomized (pre-rand) MRS 11 item total score has strong association with the primary endpoint (p-value of 0.004 in the univariate analysis and p-value of 0.003 in the multivariate analysis), whereas the pre-rand CGI-I score, the pre-rand CGI-S score, and the pre-rand GAS score have weak associations with the outcome ($0.05 \leq$ p-values ≤ 0.15). The pre-rand CGI-I score has the largest Std. Diff. of 23.6% with p-value $<$ 0.05 for the Wilcoxon assessment of imbalance. Although the random imbalance criterion in (3) does not contradict the expected balance of covariables from randomization (p-value=0.257), the possibility of random imbalance is suggested. The distribution for the pre-rand CGI-I score favors the placebo group, but the random imbalance of the pre-rand CGI-S score, the pre-rand GAS score, and the pre-rand MRS 11 item total score favor the test treatment group for better outcome.

4.1 Covariate-adjusted analyses under the MAR-like assumption

Analyses first proceed with the censoring of follow-up times for patients with premature discontinuation of their assigned treatment, and so they have the MAR-like assumption of non-informative independent censoring. The model-based variance estimator is used for hypothesis testing and to obtain confidence intervals throughout the application. The analysis results are shown in Table 3. With a Cox PH model with one explanatory variable for treatments (i.e., univariate Cox model), the unadjusted log hazard ratio (HR) for comparing test treatment versus placebo, is estimated by −0.393 with standard error (SE) of 0.161 and p-value of 0.014, indicating superiority of the test treatment. The multivariable Cox model, with the assumption of proportional hazards for treatment and all seven covariates, produces a larger estimate for the treatment effect (covariate-adjusted log HR of −0.410), a larger SE (0.164) and a slightly smaller p-value of 0.012 than the unadjusted Cox regression counterpart. When adjusting for the covariates via the NPANCOVA method, the estimated covariate-adjusted log HR is somewhat closer to the null (−0.374) than the unadjusted Cox estimates. With a slightly reduced SE (0.156), NPANCOVA produces a somewhat larger p-value (0.017). The decreased treatment effect after covariate adjustment with NPANCOVA is probably due to the random covariate imbalance favoring the test treatment group in the unadjusted analysis. Conversely, the Cox model with covariate adjustment often produces a point estimate for the treatment effect that is further from the null, mainly because it pertains to patients with the same profile of covariates in contrast to the population average nature of the unadjusted estimate or the adjusted estimate produced by the NPANCOVA method (Tangen and Koch, 2000; Jiang et al., 2008; Saville and Koch, 2013).

4.2 Sensitivity analyses with covariate adjustment

We first implement the sensitivity analysis with and without the covariate adjustment under $\theta = 1$. With this specification, the imputed data are produced from the conditional failure time distributions estimated with the censoring of the follow-up times of patients with

premature discontinuation, and they thereby have the MAR-like assumption of non-informative independent censoring. We perform the multiple imputations (MI) with $L = 50$ for the amount of missing information in this example on the basis of assessments for it in Section 4.1 of Zhao et al. (2014). Table 4 presents the covariate-adjusted (log) hazard ratios obtained from the combinations of the MI strategies and the covariate adjustment methods for the imputed data sets. An important component that differentiates various MI procedures is the survival distribution estimates, from which the conditional failure time distribution for imputation is constructed. Table 5 summarizes the methods for estimating the survival distributions, along with the other key steps and assumptions for the corresponding sensitivity analyses presented in Table 4. For the ease of comparisons, the unadjusted (log) hazard ratios were also estimated using the KMMI and the PHMI methods without covariate adjustment.

When the imputed data sets are produced by the unadjusted PHMI method under $\theta = 1$, the unadjusted (log) hazard ratio (row 4) is very similar to that obtained via the conventional unadjusted Cox model with the censoring of follow-up times for discontinued patients, mainly because both are under the MAR-like assumption of non-informative independent censoring and both have the proportional hazards assumption. With the data sets imputed by the unadjusted PHMI method, the multivariable (i.e., covariate-adjusted) Cox regression model produces a smaller treatment effect estimate and a larger SE (row 5b) than the unadjusted counterparts. Interestingly, the covariate-adjusted PHMI method (row 6b) produces a Cox model covariate-adjusted (log) hazard ratio of −0.389 that is comparable in value to the unadjusted (log) hazard ratio estimate of −0.389 from the unadjusted PHMI method (row 4). Additionally, its corresponding SE estimate is in between the SE estimates from the unadjusted and adjusted Cox regression analyses for the data imputed by the unadjusted PHMI method (row 4 and 5b). For both the unadjusted PHMI method and the covariate adjusted PHMI method, the covariate adjusted log hazard ratio estimates from NPANCOVA have somewhat smaller SE estimates than their counterparts from both the unadjusted Cox model and the corresponding multivariable Cox model.

When the data are from randomized clinical trials, one could regard the patients of each treatment group as comparable to a random sample from the study population. Therefore, it is appropriate to apply either unadjusted or covariate-adjusted analysis to the data sets imputed by the KMMI method without covariate adjustment. The unadjusted log hazard ratio from the unadjusted KMMI method (row 1) under $\theta = 1$ is closer to the null and has a somewhat larger p-value than its conventional counterpart with the use of censoring (log HR = −0.323 with p=0.044 for unadjusted KMMI versus log HR= −0.393 with p=0.014 for the conventional method), due to the tendency for non-proportional hazards for the follow-up period (i.e., much stronger effect size for the test treatment during the early part than the later part in the example). With the same set of imputed data, the estimators for a covariate-adjusted (log) hazard ratio from the multivariable Cox regression (row 3) and the NPANCOVA (row 2), produce smaller (i.e., closer to the null) treatment effects than the unadjusted Cox model estimator (row 1), which leads to larger p-values (>0.05) for both covariate-adjusted methods. Among those covariate adjustment analyses, only the NPANCOVA method generates an SE estimate smaller than that of the unadjusted Cox regression, whereas the multivariable Cox regression produces the largest SE estimates, and

hence has the largest p-value. Although the unadjusted and adjusted PHMI methods support superiority of the test treatment, the KMMI sensitivity analyses with covariate adjustment only show marginal benefits for the test treatment under $\theta = 1$.

We then conduct the covariate-adjusted sensitivity analyses, i.e., the unadjusted KMMI/unadjusted PHMI with NPANCOVA (Tables 4 and 5, row 2 and 5a), and the covariate-adjusted PHMI (Tables 4 and 5, row 6b), by varying the sensitivity parameter $\theta (=\theta_T/\theta_P)$ for the test treatment group. For sensitivity analyses in the regulatory setting, one would usually have $\theta_P = 1$ and $\theta_T > \theta_P > 1$ to penalize premature discontinuations for the test treatment. The choice of θ can be values in a range of (L, U), where $(1/U, 1/L)$ is a range of hazard ratios from previous related studies or clinical judgment for the comparison of effective medicines with placebo. Here, we set a lower bound <1 and vary the value of θ by an 0.01 increment in a range from 0.5 to 2.5, mainly because one of the three sensitivity analysis methods fails to show the superiority of the test treatment under $\theta = 1$ for this example; see Zhao et al. (2014) for additional discussion concerning the specification of the sensitivity parameter θ.

The p-values for the covariate-adjusted (log) hazard ratios from the unadjusted KMMI with NPANCOVA, the unadjusted PHMI with NPANCOVA, and the covariate-adjusted PHMI with the multivariable Cox model are plotted as functions of the sensitivity parameter θ in Figure 1. In order to have p<0.05 via the unadjusted KMMI with NPANCOVA, $\theta < 1$ is needed. This pattern of results suggests some limitation to robustness with respect to the post-discontinuation behavior of patients with the test treatment. This limitation could be due to a departure from the proportional hazards assumption for treatments as discussed in Zhao et al. (2014) in terms of the hazard ratio being further from the null in the earlier part of the follow-up period than the later part; and it could also be possibly due to covariate adjustment for random imbalances favoring the test treatment. For p<0.05 with the unadjusted PHMI with NPANCOVA and the covariate-adjusted PHMI with the multivariable Cox model, $\theta < 1.52$ is needed, suggesting better robustness to assumptions about patients with premature discontinuation of treatment for this example. Compared with the unadjusted hazard ratio estimates obtained from the unadjusted PHMI method with the specification of $\theta > 1$ (presented in Zhao et al. (2014)), the covariate-adjusted PHMI method with the multivariable Cox model produces slightly weaker results, i.e., treatment effect estimates that are closer to the null and have larger SE estimates and larger p-values.

5. Summary

Covariate adjustment can play an important role in the analysis of randomized clinical trials. In this regard, covariate adjustment may provide more powerful statistical tests (relative to their unadjusted counterparts) for the comparison between treatment groups (Koch et al., 1982, 1998). In this paper, we discussed some covariance analysis methods for time-to-event data through an example from a clinical trial for a maintenance treatment of bipolar disorder, in which substantial premature discontinuations of treatment occurred. The goal of this paper is to illustrate how to adapt the methods for covariate adjustment to the sensitivity analysis for assessing the robustness of conclusions to the management of missing information.

The multivariable Cox proportional hazards model is commonly employed for covariate adjustment in randomized studies. However, the appropriate application depends on several assumptions, such as correct model specification and proportional hazards for each variable in the model. When the proportional hazards assumption is not satisfied, the type 1 error can be inflated for the Cox model with adjustment for covariables that are related to the outcome (Jiang et al., 2008). With adjustment for covariates, the treatment parameter estimates from the Cox model are often further from the null, and the corresponding SE estimates tend to be larger than the unadjusted counterparts. Therefore, the efficiency of the null hypothesis test of no treatment effect may not be clear for covariate adjustment (Hauck et al., 1998). To implement the covariate adjustment with multivariable Cox models in the sensitivity analysis for missing data, the imputed data sets are generated and analyzed by the Cox proportional hazards model with treatment and the set of covariates to be adjusted; and this could lead to the same issues as previously noted and cause concerns for interpreting the adjusted results, especially in the regulatory setting.

The NPANCOVA method proposed by Saville and Koch (2012) has random assignment of treatments as the principal assumption and avoids the major issues associated with the multivariable Cox proportional hazards model. Unlike the multivariable Cox model, the NPANCOVA method is more likely to preserve the type 1 error under non-proportional hazards and is more robust for different model assumptions (Jiang et al., 2008; Saville and Koch, 2013). The covariate-adjusted hazard ratio produced by NPANCOVA has the interpretation of a population average treatment effect, in contrast to the subpopulation (defined by adjusted covariates) specific estimates provided by the multivariable Cox model. If adjusted covariates explain some of the variation in the response variable, the NPANCOVA method could generate more powerful statistical tests through variance reduction (Koch et al., 1982, 1998). In addition, the covariate adjustment with NPANCOVA induces equivalent comparison groups by offsetting random imbalances between treatment groups for covariables with noteworthy associations with the outcome of interest, and thereby it provides clarification of the degree to which the detected difference between randomized groups for the response variable is due to treatment rather than random imbalances for covariates. The main limitation of the NPANCOVA method is that its scope does not provide estimation for the effect of the covariates with adjustment or for the interactions of these covariates with the treatment.

For the application data set with random covariate imbalances suggestively favoring the test treatment group to have better response, the NPANCOVA covariate-adjusted sensitivity analysis with unadjusted PHMI or covariate adjusted PHMI support robustness for their unadjusted and covariate adjusted conventional counterparts with the use of censoring for this particular example. On the other hand, the NPANCOVA method with KMMI produces a covariate-adjusted log hazard ratio closer to the null with SE reduction, but a larger p-value than the counterpart produced by the unadjusted Cox regression. This limitation in robustness, as noticed with KMMI in Zhao et al. (2014), mainly seems due to a departure from the proportional hazards assumption with the hazard ratio for treatments tending to be further from the null in the earlier part of the follow-up period than the later part.

Acknowledgments

The first author gratefully acknowledges the support of Robert W. Tipping and Dr. Kenneth Koury, and thanks them for reading this paper. The authors acknowledge GlaxoSmithKline for generously providing data for the clinical trial example of the maintenance treatment for bipolar disorder. The views and opinions contained in this article shall not be construed or interpreted whether directly or indirectly to be the views or opinions of any of the officers or employees of GlaxoSmithKline Research and Development Limited or any of its affiliated companies forming part of the GlaxoSmithKline group of companies. Further, reliance on the information contained in this article is at the sole risk of the user. The information is provided “as is” without any warranty or implied term of any kind, either express or implied, including but not limited to any implied warranties or implied terms as to quality, fitness for a particular purpose, or non-infringement. All such implied terms and warranties are hereby excluded.

Funding

This paper is supported in part by NIH R01 ES021900 (H.Z.).

References

Austin, PC.; Chiu, M.; Ko, DT.; Geoeree, Ron; Tu, JV. Propensity score matching for estimating treatment effects. In: Faries, DE.; Leon, AC.; Haro, JM.; Obenchain, RL., editors. Analysis of Observational Health Care Data Using SAS. Vol. 3. SAS Institute Inc; Cary, NC: 2010. p. 51-84.

Calabrese JR, Bowden CL, Sachs G, Yatham LN, Behnke K, Mehtonen OP, Montgomery P, Ascher J, Paska W, Earl N, DeVeaugh-Geiss J. Group, Lamictal 605 Study. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. Journal of Clinical Psychiatry. 2003; 64(9):1013–1024. [PubMed: 14628976]

Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological). 1972; 34(2):187–220.

Grizzle JE, Starmer CF, Koch GG. Analysis of categorical data by linear models. Biometrics. 1969; 25:489–504. [PubMed: 5824401]

Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials. Controlled Clinical Trials. 1998; 19:249–256. [PubMed: 9620808]

Jiang H, Symanowski J, Paul S, Qu Y, Zagar A, Hong S. The type i error and power of non-parametric logrank and wilcoxon tests with adjustment for covariates – a simulation study. Statistics in Medicine. 2008; 27:5850–5860. [PubMed: 18759373]

Koch GG, Amara IA, Davis GW, Gillings DB. A review of some statistical methods for covariance analysis of categorical data. Biometrics. 1982; 38:563–595. [PubMed: 6756493]

Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine. 1998; 17:1863–1892. [PubMed: 9749453]

Lin DY, Wei LJ. The robust inference for the cox proportional hazard model. Journal of the American Statistical Association. 1989; 84(408):1074–1078.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. John Wiley & Sons, Inc; New York: 2002.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information journal. 2008; 42(4):303–319.

Moodie PF, Saville BR, Koch GG, Tangen CM. Estimating covariate-adjusted log hazard ratios for multiple time intervals in clinical trials using nonparametric randomization based ANCOVA. Statistics in Biopharmaceutical Research. 2011; 3(2):232–241.

National Research Council, Panel on Handling Missing Data in Clinical Trials. The Preventing and treatment of Missing Data in Clinical Trial. The National Academies Press; 2010. http:\\www.nap.edu\openbook.php?recordid=12955

Rubin, DB. Multiple Imputation for Nonresponse in Survey. John Wiley & Sons, Inc; New York: 1987.

Rubin DB, Schenker N. Multiple imputations in health-care database: an overview and some applications. Statistics in Medicine. 1991; 10(4):585–598. [PubMed: 2057657]

Saville BR, Koch GG. Estimating covariate-adjusted log hazard ratios in randomized clinical trials using cox proportional hazards models and nonparametric randomization based analysis of covariance. Journal of Biopharmaceutical Statistics. 2013; 23(2):477–490. [PubMed: 23437952]

Tangen CM, Koch GG. Non-parametric covariance methods for incidence density analysis of time-to-event data from a randomized clinical trial and their complementary roles to proportional hazards regression. Statistics in Medicine. 2000; 19:1039–1058. [PubMed: 10790679]

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. Journal of the American Statistical Association. 1989; 84(408): 1065–1073.

Zhang J. Sensitivity analysis of missing data: Case studies using model-based multiple imputation. Drug Information Journal. 2009; 43:475–484.

Zhao Y, Herring AH, Zhou H, Ali MW, Koch GG. A multiple imputation strategy for sensitivity analyses of time-to-event data with possibly informative censoring. Journal of Biopharmaceutical Statistics. 2014; 24(2):229–253. [PubMed: 24605967]

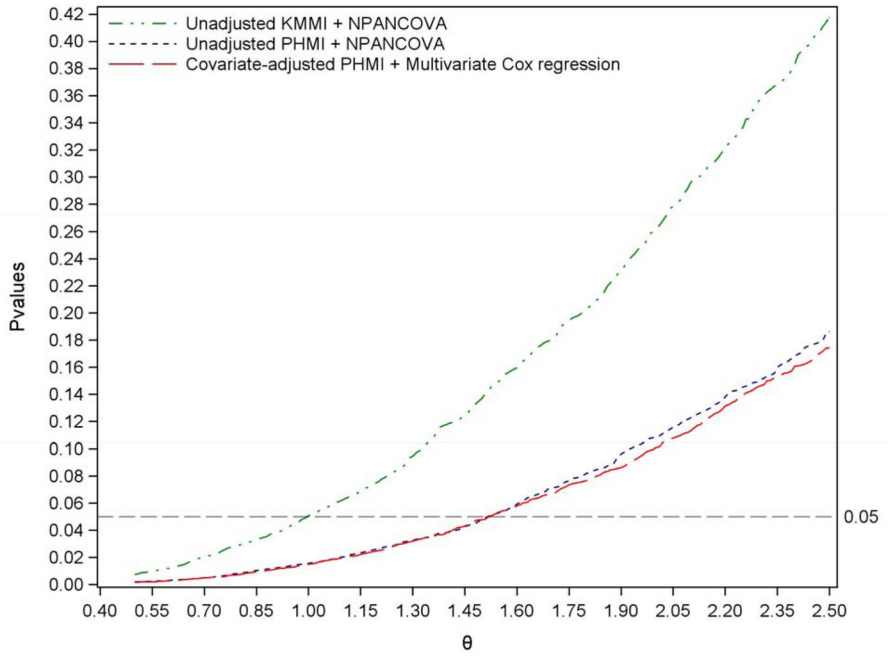

Figure 1.
Sensitivity analyses with covariate adjustment

Table 1

Distribution of patients' baseline characteristics

(N) Covariables	Overall (300) Mean	Test treatment (150) Mean (STD)	Placebo (150) Mean (STD)	Std Diff (%) (Test – Placebo)	p-values (Wilcoxon) test
Age	42.9	42.7 (11.4)	43.2 (13.3)	−5	0.550
Female (%)	49.7	45.3 (50.0)	54.0 (50.0)	−17	0.134
Pre-rand CGI-I score	1.67	1.60 (0.58)	1.74 (0.61)	−24	0.047
Pre-rand CGI-S score	2.02	1.99 (0.71)	2.06 (0.77)	−10	0.352
Pre-rand GAS score	75.1	76.0 (9.8)	74.2 (11.0)	17	0.240
Pre-rand MRS 11 item total score	1.59	1.47 (2.51)	1.72 (2.80)	−10	0.793
Pre-rand HAMD 17 item total score	5.84	5.69 (4.23)	5.99 (4.17)	−7	0.422

Table 2

Associations of patients' baseline characteristics and the primary outcome (assessed with Cox model)

Covariables	Univariate analysis		Multivariate Analysis	
	Coefficient (SE)	p-value	Coefficient (SE)	p-value
Age	−0.001 (0.007)	0.932	0.001 (0.007)	0.890
Female (proportion)	0.046 (0.162)	0.776	−0.004 (0.164)	0.982
Pre-rand CGI-I score	0.055 (0.128)	0.669	−0.380 (0.219)	0.082
Pre-rand CGI-S score	0.149 (0.104)	0.153	0.260 (0.165)	0.115
Pre-rand GAS score	−0.013 (0.008)	0.093	−0.015 (0.011)	0.151
Pre-rand MRS 11 item total score	0.073 (0.025)	0.004	0.077 (0.026)	0.003
Pre-rand HAMD 17 item total score	0.023 (0.019)	0.239	0.003 (0.025)	0.901

Table 3

Covariate-adjusted analyses for treatment effects under the MAR-like assumption

Method	Parameter (SE)	HR (95% CI)	p-values
Univariate (unadjusted) Cox model	−0.393 (0.161)	0.675 (0.493, 0.925)	0.014
Multivariable Cox model (adjusted)	−0.410 (0.164)	0.664 (0.482, 0.915)	0.012
NPANCOVA (adjusted)	−0.374 (0.156)	0.688 (0.507, 0.935)	0.017

Table 4

Sensitivity analysis with specification of $\theta = 1$

Imputation Method	Analysis Method	Parameter	StdErr	HR	95% CI for HR	*Pr* > |*t*|
1. Unadjusted KMMI	Univariate (unadjusted) Cox model	−0.323	0.160	0.724	(0.530, 0.991)	0.044
2. Unadjusted KMMI	Non-parametric ANCOVA (adjusted)	−0.309	0.158	0.734	(0.538, 1.001)	0.051
3. Unadjusted KMMI	Multivariable Cox model (adjusted)	−0.306	0.163	0.737	(0.536, 1.014)	0.061
4. Unadjusted PHMI	Univariate (unadjusted) Cox model	−0.389	0.158	0.678	(0.497, 0.925)	0.014
5a. Unadjusted PHMI	Non-parametric ANCOVA (adjusted)	−0.378	0.156	0.685	(0.504, 0.931)	0.016
5b. Unadjusted PHMI	Multivariable Cox model (adjusted)	−0.373	0.161	0.689	(0.502, 0.945)	0.021
6a. Covariate-adjusted PHMI	Non-parametric ANCOVA (adjusted)	−0.390	0.155	0.677	(0.500, 0.917)	0.012
6b. Covariate-adjusted PHMI	Multivariable Cox model (adjusted)	−0.389	0.159	0.678	(0.496, 0.927)	0.015

Table 5

Key steps and assumptions in the performance of sensitivity analyses under $\theta = 1$

Multiple imputation strategy		Analysis for imputed data	
Methods for obtaining survivor estimates	**PH assumption for survival distributions**	**Analysis method**	**PH assumption for covariates**
1. KM estimator for individual groups	Not required	Univariate Cox model with treatment as the only explanatory variable	Not required
2. KM estimator for individual groups	Not required	Non-parametric ANCOVA	Not required
3. KM estimator for individual groups	Not required	Multivariable Cox model with treatment and the seven covariables	Required
4. Breslow estimator from univariate Cox model with treatment as explanatory variable	Required for treatment groups	Univariate Cox model with treatment as explanatory variable	Not required
5. Breslow estimator from univariate Cox model with treatment as explanatory variable	Required for treatment groups	a. Nonparametric ANCOVA	Not required
		b. Multivariable Cox model with treatment and the seven covariables	Required
6. Breslow estimator from multivariable Cox model with treatment and covariables	Required for treatment groups and covariates	a. Nonparametric ANCOVA	Not required
		b. Multivariable Cox model with treatment and the seven covariables	Required