Harvard University
Harvard University Biostatistics Working Paper Series

Year 2016 *Paper* 209

Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study

Fei Jiang[*] Lu Tian[†] Haoda Fu[‡]
Takahiro Hasegawa[**] Marc Alan Pfeffer[††] L. J. Wei[‡‡]

[*]The University of Hong Kong, feijiang@hku.hk
[†]Stanford University School of Medicine, lutian@stanford.edu
[‡]Eli Lilly and Company, fu_haoda@lilly.com
[**]Shionogi & Co., Ltd, duo@mars.dti.ne.jp
[††]Brigham and Women's Hospital, mpfeffer@rics.bwh.harvard.edu
[‡‡]Harvard University, wei@hsph.harvard.edu

http://biostats.bepress.com/harvardbiostat/paper209

Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study

Fei Jiang, Lu Tian, Haoda Fu, Takahiro Hasegawa, Marc Alan Pfeffer, and L. J. Wei

Abstract

In comparing two treatments via a randomized clinical trial, the analysis of covariance technique is often utilized to estimate an overall treatment effect. The ANCOVA is generally perceived as a more efficient procedure than its simple two sample estima- tion counterpart. Unfortunately when the ANCOVA model is not correctly specified, the resulting estimator is generally not consistent especially when the model is nonlin- ear. Recently various nonparametric alternatives, such as the augmentation methods, to ANCOVA have been proposed to estimate the treatment effect by adjusting the covariates. However, the properties of these alternatives have not been studied in the presence of treatment allocation imbalance. In this paper, we take a different approach to explore how to improve the precision of the naive two-sample estimate even when the observed distributions of baseline covariates between two groups are dissimilar.

Specifically, we derive a bias-adjusted estimation procedure constructed from a condi- tional inference principle via relevant ancillary statistics from the observed covariates. This estimator is shown to be asymptotically equivalent to an augmentation estimator under the conditional setting. We utilize the data from a clinical trial for evaluating a combination treatment of cardiovascular diseases to illustrate our findings.

Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study

Fei Jiang
Department of Statistics & Actuarial Science,
University of Hongkong, Hongkong
(*feijiang@hku.hk*)
Lu Tian*
Department of Biomedical Data Science,
Stanford University, Stanford, California 94305, U.S.A.
(*lutian@stanford.edu*)
Haoda Fu
Eli Lilly and Company
Lilly Corporate Center, Indianapolis, IN 46285, U.S.A.
(*haoda-fu@lilly.com*)
Takahiro Hasegawa
Shionogi & Co., Ltd, Osaka, Japan
(*duo@mars.dti.ne.jp*)
Marc Alan Pfeffer
Brigham and Women's Hospital Department of Medicine
Boston, Massachusetts 02115, U.S.A.
(*mpfeffer@rics.bwh.harvard.edu*)
L.J. Wei
Department of Biostatistics, Harvard University,
Boston, Massachusetts 02115, U.S.A.
(*wei@sdac.harvard.edu*)

Abstract

In comparing two treatments via a randomized clinical trial, the analysis of covariance technique is often utilized to estimate an overall treatment effect. The ANCOVA is generally perceived as a more efficient procedure than its simple two sample estimation counterpart. Unfortunately when the ANCOVA model is not correctly specified, the resulting estimator generally is not consistent especially when the model is nonlinear. Recently various nonparametric alternatives, such as the augmentation methods, to ANCOVA have been proposed to estimate the treatment effect by adjusting the covariates. However, the properties of these alternatives have not been studied in the

*co-first author

presence of treatment allocation imbalance. In this paper, we take a different approach to explore how to improve the precision of the naive two-sample estimate even when the observed distributions of baseline covariates between two groups are dissimilar. Specifically, we derive a bias-adjusted estimation procedure constructed from a conditional inference principle via relevant ancillary statistics from the observed covariates. This estimator is shown to be asymptotically equivalent to an augmentation estimator under the conditional setting. We utilize the data from a clinical trial for evaluating a combination treatment of cardiovascular diseases to illustrate our findings.

Keywords: Ancillary statistic; Augmentation estimation procedure; Conditional inference; Stratified analysis

1 Introduction

In comparing two treatment groups, let θ be the parameter of interest for quantifying the between-group difference with respect to the study endpoint. For example, let Y be the outcome variable, Z be the binary treatment indicator, $\mu_0 = E(Y|Z=0), \mu_1 = E(Y|Z=1)$, and $\theta = \mu_1 - \mu_0$. Let $\hat{\theta}$ be the corresponding two-sample estimator with the data from a comparative, randomized clinical trial with a $M:1$ treatment allocation rule (Group 0 vs. 1). If Y is a binary outcome, θ may be the risk or odds ratio. In general, with a large sample size, the distribution of $\hat{\theta}$ is approximately normal with mean θ. Inferences about θ can be made accordingly. When the patient's potentially predictive baseline covariate vector $\mathbf{X}$ is available, one may utilize an analysis of covariance (ANCOVA) procedure to estimate θ. However, when the ANCOVA model is non-linear (e.g., a logistic or proportional hazard model) and not correctly specified, the resulting estimator for the treatment effect is not consistent to θ (Gail et al., 1984; Lin & Wei, 1989; Struthers & Kalbfleisch, 1986). For this case, various robust, nonparametric adjust estimation procedures for θ have been proposed, which are well summarized in a recent paper by Rosenblum & van der Laan (2010). For instance, an argumentation estimation procedure with covariate adjustment provides a consistent estimator for θ (Leon et al., 2003; Tsiatis, 2006; Tsiatis et al., 2008; Lu & Tsiatis, 2008; Zhang et al., 2008; Gilbert et al., 2009; Zhang & Gilbert, 2010; Tian et al., 2012). Such an estimator, say, $\hat{\theta}_{aug}$, is or asymptotically equivalent to a linear combination of $\hat{\theta}$ and $\widehat{\Delta}_{\mathbf{X}} = \bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0$, where $\bar{\mathbf{X}}_k$ is the sample mean of the covariate vectors or a transformation thereof, for treatment $k, k = 0, 1$. The distribution of $\hat{\theta}_{aug}$ is also approximately normal with mean θ. The standard error estimate for $\hat{\theta}_{aug}$ can be substantially smaller than that based on $\hat{\theta}$ when the augmented covariates are highly correlated with the response endpoint. Unlike the ANCOVA, the augmentation method is a model-free technique. Note that the stochastic

properties of the above estimators were studied only under an unconditional setting in the literature, that is, with the study size n, their sample space is generated by all possible realizations of a random sample consisting of n independent, identically distributed copies of $(Y, Z, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$. Also note that under the unconditional setting, the naive estimator $\hat{\theta}$ is consistent.

Another important goal of utilizing the covariate-adjustment technique for estimating the treatment difference is to reduce bias of the simple estimator $\hat{\theta}$ when, by random chance, the observed distributions of the covariate vectors are dissimilar between two comparative groups. For this case, $\hat{\theta}$ can be severely biased. It is not clear, however, that the above robust alternatives would be better than $\hat{\theta}$ in the presence of covariate imbalance. To quantify the potential bias from $\hat{\theta}$, one may consider an appropriate sample space other than the aforementioned unconditional one. For instance, if the ANCOVA model is correctly specified, conditional on all observed covariate vectors, an asymptotically unbiased estimator for θ can be constructed accordingly, but the naive estimate $\hat{\theta}$ can be severely biased under this conditional setting. Note that the study subjects' covariates are ancillary statistics, that is, they are not directly related to the treatment difference θ. In general, analyzing data under a conditional inference principle on certain ancillary statistics makes the resulting inference more relevant to the observed data (Cox, 1958; Cox & Hinkley, 1979; Fraser & McDunnough, 1980; Berger et al., 1988; Casella, 1992; Fraser et al., 2004; Ghosh et al., 2010). Under the ANCOVA setting, to study the stochastic behavior of an estimator for θ, the sample space consists of all realizations from a random sampling scheme, but the covariate vectors of each realization would be the same as the observed counterpart so that the resulting realized estimate is generated under the most "similar" experimental condition as the observed profile of baseline covariates. It may be desirable to consider such a fine level of the conditional setting to obtain consistent estimators for θ in the presence of covariate imbalance. Unfortunately, the ANCOVA model is likely misspecified in practice.

The choice of the conditioning ancillary statistic is not unique (Basu, 1959; Cox, 1971; Ghosh et al., 2010). For the present case, instead of conditioning on the entire set of observed covariates, a relevant ancillary statistic one may consider to study the stochastic behavior of estimators for θ would be the aforementioned $\widehat{\Delta}_{\mathbf{X}} = \bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0$, which is a natural, and commonly used summary measure of covariate-imbalance in clinical studies (Pocock et al., 2002). That is, we only consider the realization of the sample space generated from the randomized clinical trial setting, whose imbalance measured by the two-sample covariate mean difference is identical to the observed counterpart. In this paper, we show that based on this conditional inference principle, a bias-adjusted estimator $\hat{\theta}_{adj}$ from $\hat{\theta}$ is asymptotically equivalent to the aforementioned augmentation procedure derived from the uncondtional setting. We used the data from a comparative clinical trial to evaluate treatments for cardiovascular diseases to illustrate our findings. Furthermore, a numerical

study is conducted to examine the performance of $\hat{\theta}_{adj}$. We find via this study that if the covariates of the ancillary statistics are highly correlated with the outcome variable and/or the treatment allocation proportions, $\hat{\theta}_{adj}$ can be substantially better than the two sample estimator $\hat{\theta}$.

2 The distributions of $\hat{\theta}$ conditioning on $\widehat{\Delta}_{\mathbf{X}}$ and a bias-adjusted estimator $\hat{\theta}_{adj}$

Let $\theta = g(\mu_0, \mu_1)$, where g is a smooth function. Then $\hat{\theta} = g(\hat{\mu}_0, \hat{\mu}_1)$ is the two sample naive estimator for θ. Under the random treatment assignments, $\hat{\theta} - \theta$ and $\widehat{\Delta}_{\mathbf{X}}$ are approximately normal with mean 0 and covariance matrix

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{12} & \hat{\Sigma}_{22} \end{pmatrix},$$

where

$$\begin{aligned} \hat{\Sigma}_{11} &\approx \dot{g}_1^2(\mu_0, \mu_1)\mathrm{var}(\hat{\mu}_0) + \dot{g}_2^2(\mu_0, \mu_1)\mathrm{var}(\hat{\mu}_1), \\ \hat{\Sigma}_{12} &\approx \dot{g}_2(\mu_0, \mu_1)\mathrm{cov}(\hat{\mu}_1, \bar{\mathbf{X}}_1) - \dot{g}_1(\mu_0, \mu_1)\mathrm{cov}(\hat{\mu}_0, \bar{\mathbf{X}}_0), \quad \text{and} \\ \hat{\Sigma}_{22} &\approx \mathrm{var}(\bar{\mathbf{X}}_0) + \mathrm{var}(\bar{\mathbf{X}}_1), \end{aligned}$$

are the estimated variance of $\hat{\theta} - \theta$, the estimated covariance matrix between $\widehat{\Delta}_{\mathbf{X}}$ and $\hat{\theta} - \theta$ and the estimated covariance matrix of $\widehat{\Delta}_{\mathbf{X}}$, respectively. Here $\dot{g}_1$ and $\dot{g}_2$ are the partial derivatives of g with respect to the first and second argument. Now, let $\mathbf{d}$ be the observed value of $\widehat{\Delta}_{\mathbf{X}}$. Then for large n, the distribution of $\hat{\theta} - \theta$ given $\widehat{\Delta}_{\mathbf{X}} = \mathbf{d}$ is approximately normal with mean $\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\mathbf{d}$, and variance $\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$. It follows that, when $\mathbf{d}$ is not zero, $\hat{\theta}$ is not consistent under this conditional argument and a bias-adjusted estimator for θ is

$$\hat{\theta}_{adj} = \hat{\theta} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\mathbf{d}.$$

The justification of this conditional distribution approximation is not straightforward and the details as well as the sufficient conditions are given in Appendix A. The conditions to ensure a Gaussian approximation to the conditional distribution are not too stringent. They are satisfied, for instance, when $\hat{\theta}$ is a regular estimator with a limiting Gaussian distribution and the covariates are bounded with a non-singular variance-covariance matrix.

For example, when θ is the log-transformed odds ratio (OR), i.e., $g(\mu_0, \mu_1) = \log\left\{\frac{\mu_1(1-\mu_0)}{\mu_0(1-\mu_1)}\right\}$,

$$\begin{aligned}
\hat{\theta}_{adj} &= \log\left\{\frac{\hat{\mu}_1(1-\hat{\mu}_0)}{\hat{\mu}_0(1-\hat{\mu}_1)}\right\} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\widehat{\Delta}_{\mathbf{X}}, \\
\hat{\Sigma}_{11} &= \frac{1}{n_1\hat{\mu}_1} + \frac{1}{n_1(1-\hat{\mu}_1)} + \frac{1}{n_0\hat{\mu}_0} + \frac{1}{n_0(1-\hat{\mu}_0)}, \\
\hat{\Sigma}_{12} &= \frac{\hat{\Sigma}_{121}}{n_1\hat{\mu}_1(1-\hat{\mu}_1)} + \frac{\hat{\Sigma}_{120}}{n_0\hat{\mu}_0(1-\hat{\mu}_0)}, \quad \text{and} \\
\hat{\Sigma}_{22} &= \frac{\hat{\Sigma}_{221}}{n_1} + \frac{\hat{\Sigma}_{220}}{n_0},
\end{aligned}$$

where n_k, $\hat{\Sigma}_{12k}$ and $\hat{\Sigma}_{22k}$ are the sample size, empirical covariance between Y and $\mathbf{X}$ and the variance-covariance matrix of $\mathbf{X}$ in arm k, $k = 0, 1$, respectively.

Note that $\hat{\theta}_{adj}$ is equivalent or asymptotically equivalent to augmentation estimators (Tsiatis et al., 2008; Tian et al., 2012). The justification is given in Appendix B.

3 Example and a numerical study

In this section, we used the data from a cardiovascular trial: "Valsartan in acute myocardial infarction" (VALIANT) study (Pfeffer et al., 2003) to illustrate our findings. The study patients were equally randomized to three arms: ARB valsartan, captopril and a combination of these two drugs. Here, we consider a binary outcome as the endpoint, which indicates whether the patient had hospitalization/death by Month 18. Since there was no difference between two mono-therapies with respect to this outcome, we combined the data from two mono-therapy groups to evaluate the combo-treatment effect. The study enrolled a total of 14,703 patients. The observed event rates for mono- and combo are 0.58 and 0.57, indicating that there was no benefit from the combo with respect to this outcome. On the other hand, with the data from 302 patients in Australia, the mono-therapy somehow appears to be statistically significantly better than its combo counterpart based on the simple two sample estimate (the observed event rates for combo and mono are 0.80 and 0.67). Now, let θ be the log OR, and $\hat{\theta}$ be its naive estimate. The point estimate of OR (combo vs. mono), i.e., $\exp(\hat{\theta})$ and 0.95 confidence interval are 1.99 and (1.12, 3.51), respectively. Note that Australia was the only country among 24 countries participated in the VALIANT study, whose patients appear to have better outcomes for the mono-therapy. It is not clear whether the Australian patients were quite different from those from the rest of world to have such a discrepancy on the treatment effect profile.

To explore this further for the Australia study, we found that there was treatment allocation imbalance between these two treatment groups with respect to, for example, the patients binary pre-existing diabetes status (DIAS) and baseline heart rate (HR). In Figure

1, we show the fitted curves stratified by DIAS via two logistic regression models with the treatment assignment being the outcome and standerdized HR, HR^2 and HR^3 as the independent variables. If the randomization treatment allocation scheme were working for the Australia study, these two curves would be flat around 2/3. Figure 1 indicates that there was indeed nontrivial treatment allocation imbalance between the mono and combo groups. Now, let $\hat{\theta}_{adj}$ be the biased-adjusted estimate for the log OR. The corresponding bias-adjusted estimator of OR, i.e., $\exp(\hat{\theta}_{adj})$ and the 0.95 confidence interval conditional on DIAS, HR, HR^2 and HR^3 are 1.68 and (0.95, 2.94), respectively. Note that the confidence interval contains the null value of 1. Also note that one of the reasons we considered the HR variable up to the third order for the conditioning inference is that most distributions can be characterized with their first three moments. This conditioning setting would be approximately the same as that with the entire distribution of HR.

To explore whether increasing the degree of correlation between the outcome and the covariates would increase the precision of $\hat{\theta}_{adj}$, we considered three cases of correlation profiles between the outcome and DIAS/HR. In Figure 2(a), we show the fitted curves stratified by DIAS via two logistic regression models with the binary outcome and independent variables of HR, HR^2 and HR^3 using the entire data over three arms from VALIANT. Note that DIAS seems correlated with the outcome, but not the baseline HR. We then used these two regression models and the curves in Figure 1 to generate the outcomes and the treatment assignments for the Australian patients with their observed covariates. With 1000 simulated set of realizations, in Figure 2(b), we present the empirical density function estimates for the naive and bias-adjusted estimators of OR. It appears that the improvement from the adjusted estimator over the naive estimator is modest for this scenario. For this case, the average absolute difference between the true parameter value and the estimator is 0.99 for the naive estimator and 0.68 for the bias-adjusted estimator.

On the other hand, when we increase the degree of correlation between the outcome and the covariates (See Figures 3 and 4), $\exp(\hat{\theta}_{adj})$ performs much better than $\exp(\hat{\theta})$ with respect to the estimation precision. For instance, the empirical absolute biases for $\exp(\hat{\theta})$ and $\exp(\hat{\theta}_{adj})$ are 0.15 and 0.01 for Figure 3(b) and are 0.29 and 0.003 for Figure 4(b). Similar phenomena has also been observed when the degree of treatment allocation imbalance increases with respect to the covariates, the precision of the bias-adjusted estimator increases over the naive one.

4 Remarks

In this paper, we take a different angle to explore how to improve the precision of the naive two sample estimator $\hat{\theta}$ for the treatment effect with the patients baseline covariates. Conceptually our approach by conditioning on the ancillary statistics is similar to

that of ANCOVA, but without heavy modeling involved. The resulting estimate is asymptotically equivalent to a certain augmentation estimation counterpart, which was derived under the unconditional setting from a semi-parametric efficiency argument. Like other covariate-adjusted estimation procedures, the proper choice of the covariates to construct the ancillary statistics for our proposal is crucial. The precision gain can be substantial if the covariates included in the analysis are highly correlated with the outcome and/or the treatment allocation proportions. Under the unconditional setting, the simple two sample estimator $\hat{\theta}$ and any augmentation estimators are consistent, Tian et al. (2012) proposed a sequential procedure to select an optimal set of covariates for the augmentation method. However, it is not clear how to generalize their method to the current conditional setting. Further research along this line is needed.

Stratified analysis can be regarded as a special case of the covariate-adjusted procedure. On the other hand, due to its discrete nature of possible values of the covariates, using the present conditioning approach, one may consider the ancillary statistic consisting of the entire observed covariate vectors for stratified analysis (Tian et al., 2016). For the general case when some of the covariates are continuous, such a fine level of conditioning would be difficult, if not impossible to implement.

Lu & Tsiatis (2008) and Tian et al. (2012) discussed the augmentation method with covariates when the outcome is an event time observation. It is straightforward to show that our bias-adjusted estimator conditional on the empirical averages covariate imbalance measures is asymptotically equivalent to an augmentation estimator with censored observations via the justification in Appendix B. Note that in this case, the conventional ANCOVA with the Cox model may result in a hazard ratio estimate for the group contrast measure, which is difficult to interpret clinically.

Figure 1. The treatment allocation proportions to mono-therapy arm: solid line is for DIAS = 1; dashed line is for DIAS = 0.

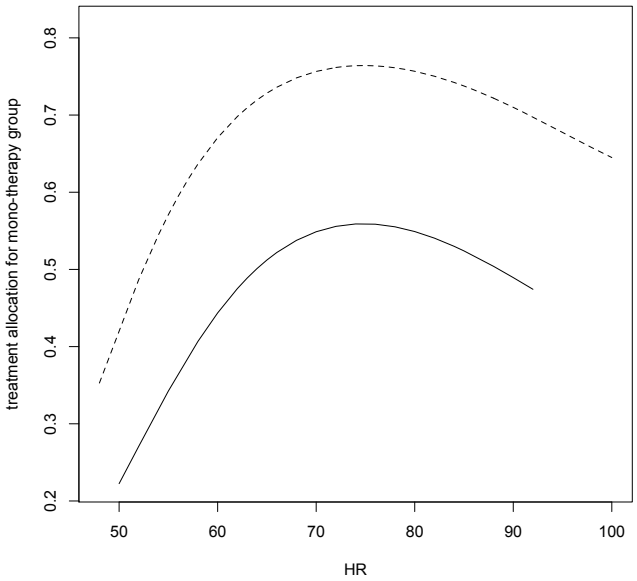


Figure 2(a). The estimated event rate curves with respect to diabetes status and heart rate at the baseline with data from the VILIANT study

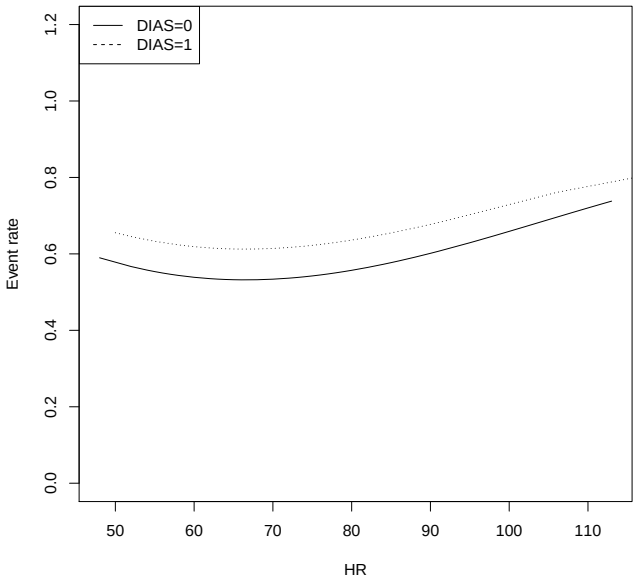


Figure 2(b). The empirical density functions for $\hat{\theta}$ and $\hat{\theta}_{adj}$

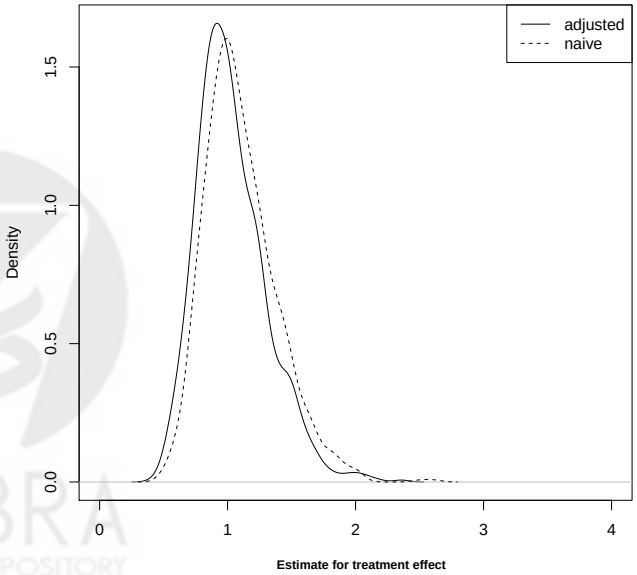


Figure 3(a). The event rate curves with respect to diabetes status and heart rate at the baseline

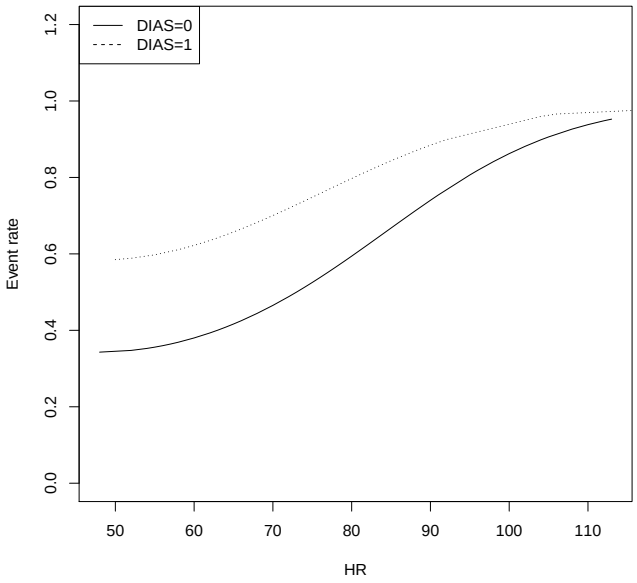


Figure 3(b). The empirical density functions for $\hat{\theta}$ and $\hat{\theta}_{adj}$

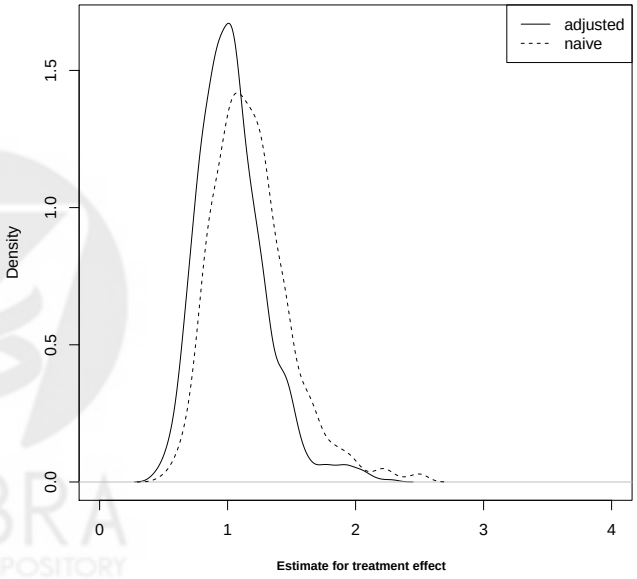


Figure 4(a). The event rate curves with respect to diabetes status and heart rate at the baseline

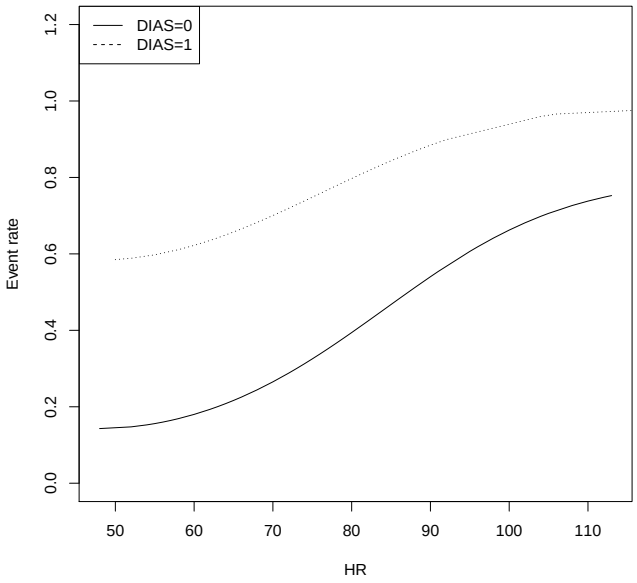
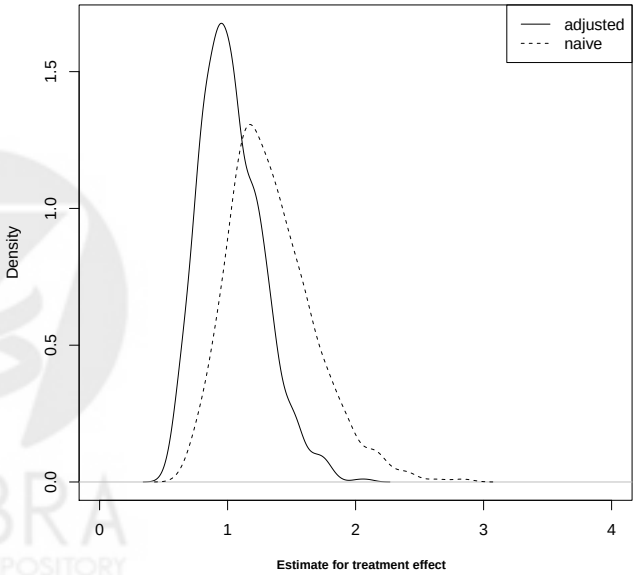


Figure 4(b). The empirical density functions for $\hat{\theta}$ and $\hat{\theta}_{adj}$



References

Basu, D. (1959). The family of ancillary statistics. *Sankhyā: The Indian Journal of Statistics* , 247–256.

Berger, J. O., Wolpert, R. L., Bayarri, M., DeGroot, M., Hill, B. M., Lane, D. A. & LeCam, L. (1988). The likelihood principle. *Lecture notes-Monograph series* **6**, iii–199.

Casella, G. (1992). Conditional inference from confidence sets. *Lecture Notes-Monograph Series* , 1–12.

Cox, D. (1971). The choice between alternative ancillary statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* , 251–255.

Cox, D. R. (1958). Some problems connected with statistical inference. *The Annals of Mathematical Statistics* , 357–372.

Cox, D. R. & Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.

Fraser, D. & McDunnough, P. (1980). Some remarks on conditional and unconditional inference for location-scale models. *Statistische Hefte* **21**, 224–231.

Fraser, D. et al. (2004). Ancillaries and conditional inference. *Statistical Science* **19**, 333–369.

Gail, M. H., Wieand, S. & Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431–444.

Ghosh, M., Reid, N. & Fraser, D. (2010). Ancillary statistics: a review. *Statistica Sinica* **20**, 1309.

Gilbert, P. B., Sato, A., Sun, X. & Mehrotra, D. V. (2009). Efficient and robust method for comparing the immunogenicity of candidate vaccines in randomized clinical trials. *Vaccine* **27**, 396–401.

Leon, S., Tsiatis, A. A. & Davidian, M. (2003). Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics* **59**, 1046–1055.

Lin, D. Y. & Wei, L.-J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American statistical Association* **84**, 1074–1078.

Lu, X. & Tsiatis, A. A. (2008). Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika* **95**, 679–694.

Pfeffer, M. A., Swedberg, K., Granger, C. B., Held, P., McMurray, J. J., Michelson, E. L., Olofsson, B., Östergren, J., Yusuf, S., Investigators, C., Committees et al. (2003). Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the charm-overall programme. *The Lancet* **362**, 759–766.

Pocock, S. J., Assmann, S. E., Enos, L. E. & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practiceand problems. *Statistics in medicine* **21**, 2917–2930.

Rosenblum, M. & van der Laan, M. J. (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The international journal of biostatistics* **6**.

Steck, G. (1957). *Limit Theorems for Conditional Distributions, by George P. Steck*. California: University of California press.

Struthers, C. A. & Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–369.

Tian, L., Cai, T., Zhao, L. & Wei, L.-J. (2012). On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial. *Biostatistics* **13**, 256–273.

Tian, L., Jiang, F., , Hasegawa, T., Uno, H. & Wei, L. (2016). Moving beyond the conventional stratified analysis to estimate an overall treatment efficacy with the data from a comparative randomized clinical study. Submitted to Biostatistics.

Tsiatis, A. A. (2006). Information-based monitoring of clinical trials. *Statistics in medicine* **25**, 3236–3244.

Tsiatis, A. A., Davidian, M., Zhang, M. & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in medicine* **27**, 4658–4677.

Zhang, M. & Gilbert, P. B. (2010). Increasing the efficiency of prevention trials by incorporating baseline covariates. *Statistical communications in infectious diseases* **2**.

Zhang, M., Tsiatis, A. A. & Davidian, M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics* **64**, 707–715.

Appendix A. Asymptotic properties of bias-adjusted estimator

Let $(Y_i, Z_i, \mathbf{X}_i^{\mathrm{T}})^{\mathrm{T}}$, $i = 1, \ldots, n$, be the independent identically distributed (i.i.d.) copies of $(Y, Z, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$. In this Appendix, we will drive the limiting distribution of

$$n^{1/2}(\hat{\theta} - \theta)$$

given $\widehat{\Delta}_{\mathbf{X}}$ under the following three conditions that

(A1) $\mathrm{cov}(Y, \mathbf{X})$ is a finite, non-degenerate matrix.

(A2) The characteristic function of $\mathbf{X}$ is integrable.

(A3) $\hat{\theta}$ is a regular estimator for θ, i.e.,

$$\hat{\theta} - \theta = n^{-1} \sum_{i=1}^{n} U_i + \xi_\theta,$$

where

$$U_i = \dot{g}_2(\mu_0, \mu_1) \frac{Z_i(Y_i - \mu_1)}{\pi} + \dot{g}_1(\mu_0, \mu_1) \frac{(1 - Z_i)(Y_i - \mu_0)}{(1 - \pi)}, i = 1, \cdots, n,$$

are i.i.d. random variables, $\pi = \mathrm{pr}(Z = 1) = 1/(M + 1)$, and $\xi_\theta = o_{a.s.}(n^{-1/2})$.

Under Condition (A3),

$$\begin{pmatrix} \hat{\theta} - \theta \\ \widehat{\Delta}_{\mathbf{X}} - \Delta_{\mathbf{X}} \end{pmatrix} = n^{-1} \sum_{i=1}^{n} \begin{pmatrix} U_i \\ \mathbf{V}_i \end{pmatrix} + \begin{pmatrix} \xi_\theta \\ \boldsymbol{\xi}_{\mathbf{X}} \end{pmatrix}, \tag{1}$$

where $\mathbf{V}_i = \pi^{-1} Z_i(\mathbf{X}_i - \boldsymbol{\tau}) + (1 - \pi)^{-1}(1 - Z_i)(\mathbf{X}_i - \boldsymbol{\tau})$, $\boldsymbol{\tau} = E(\mathbf{X})$ and $\boldsymbol{\xi}_{\mathbf{X}} = o_{a.s.}(n^{-1/2})$. Let $\mathcal{U}_n = n^{-1/2} \sum_{i=1}^{n} U_i$ and $\mathcal{V}_n = n^{-1/2} \sum_{i=1}^{n} \mathbf{V}_i$. Then $(\mathcal{U}_n, \mathcal{V}_n^{\mathrm{T}})^{\mathrm{T}}$ converges weakly to $(\mathcal{U}, \mathcal{V}^{\mathrm{T}})^{\mathrm{T}}$, a Gaussian vector with mean $\mathbf{0}$ and a finite covariate matrix $n\Sigma$, where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}.$$

Here

$$\begin{aligned} \Sigma_{11} &= n^{-1} \dot{g}_1^2(\mu_0, \mu_1) \frac{\mathrm{var}(Y | Z = 1)}{\pi} + n^{-1} \dot{g}_2^2(\mu_0, \mu_1) \frac{\mathrm{var}(Y | Z = 0)}{1 - \pi}, \\ \Sigma_{12} &= n^{-1} \dot{g}_1(\mu_0, \mu_1) \frac{\mathrm{cov}(Y, \mathbf{X} | Z = 1)}{\pi} - n^{-1} \dot{g}_2(\mu_0, \mu_1) \frac{\mathrm{cov}(Y, \mathbf{X} | Z = 0)}{1 - \pi}, \quad \text{and} \\ \Sigma_{22} &= n^{-1} \frac{\mathrm{var}(\mathbf{X})}{\pi(1 - \pi)}. \end{aligned}$$

Now, let $\{\mathbf{v}_n \in \mathbf{A}_n\}$ be any sequence of vectors such that $\mathbf{v}_n \to \mathbf{v}_0$, a constant vector, as $n \to \infty$, where $\mathbf{A}_n$ is the support of $\mathcal{V}_n$. It follows from Steck (1957) that under Conditions (A1) and (A2),

$$\sup_u |F_n^{\mathbf{v}_n}(u) - F^{\mathbf{v}_0}(u)| = o_{a.s.}(1), \tag{2}$$

where $F_n^{\mathbf{v}}(u)$ is the cumulative distribution function of the conditional distribution of $\mathcal{U}_n$ given $\mathcal{V}_n = \mathbf{v}$, and $F^{\mathbf{v}}(u)$ is the cumulative distribution function of the conditional Gaussian distribution of $\mathcal{U}$ given $\mathcal{V} = \mathbf{v}$.

Let $\mathbf{B}_n$ be the support of $n^{1/2}\widehat{\Delta}_{\mathbf{X}}$. For any sequence of vectors $\boldsymbol{\delta}_n \in \mathbf{B}_n$, such that $\boldsymbol{\delta}_n - \boldsymbol{\delta}_0 = o(1)$, $\tilde{\boldsymbol{\delta}}_n$ also converges to $\boldsymbol{\delta}_0$, as $n \to \infty$, where $\tilde{\boldsymbol{\delta}}_n = \boldsymbol{\delta}_n - \xi_{\mathbf{X}} \in \mathbf{A}_n$. Therefore,

$$\begin{aligned} & \Pr\{n^{1/2}(\hat{\theta} - \theta) \le u | n^{1/2}\widehat{\Delta}_{\mathbf{X}} = \boldsymbol{\delta}_n\} \\ = \; & \Pr(\mathcal{U}_n \le u - n^{1/2}\xi_\theta | \mathcal{V}_n = \tilde{\boldsymbol{\delta}}_n) + o_{a.s.}(1) \\ = \; & F_n^{\boldsymbol{\delta}_0}(u - n^{1/2}\xi_\theta) + o_{a.s.}(1) \\ = \; & F^{\boldsymbol{\delta}_0}(u) + o_{a.s.}(1). \end{aligned} \tag{3}$$

Note that the first equality is a direct consequence of (1), and the last equality is implied by (2) and the fact that $F^{\boldsymbol{\delta}}(u)$ is uniform continuous in u.

Now, let $\boldsymbol{\delta}_n = n^{1/2}\mathbf{d}$. Since $F^{\boldsymbol{\delta}_0}(\cdot)$ is a conditional Gaussian distribution function with mean $\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\delta}_0$, (3) implies $n^{1/2}(\hat{\theta} - \theta)$ given $n^{1/2}\widehat{\Delta}_{\mathbf{X}} = \boldsymbol{\delta}_n$ converges to a conditional Gaussian distribution with mean $n^{1/2}\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\delta}_0$ almost surely. Since $\boldsymbol{\delta}_0 - n^{1/2}\mathbf{d} = o(1)$ and $\hat{\Sigma}_{ij} - \Sigma_{ij} = o_{a.s.}(1)$, the bias-adjusted estimator $\hat{\theta} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\mathbf{d}$ is an asymptotically unbiased estimator for θ under the conditional setting with asymptotic variance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Appendix B. Equivalence between $\hat{\theta}_{aug}$ and $\hat{\theta}_{adj}$

Firstly, the efficiency-augmented estimator for $\theta = g(\mu_0, \mu_1)$ studied by Tsiatis et al. (2008) and Zhang et al. (2008) is given by

$$\hat{\theta}_{aug} = g(\mu_0^\dagger, \mu_1^\dagger),$$

where

$$\mu_1^\dagger = \hat{\mu}_1 - \sum_{i=1}^{n_1}(1-\pi)\left\{n_1^{-1}\hat{a}_1(\mathbf{X}_{1i}) - n_0^{-1}\hat{a}_1(\mathbf{X}_{0i})\right\},$$

$$\mu_0^\dagger = \hat{\mu}_0 - \sum_{i=1}^{n_0}\pi\left\{n_0^{-1}\hat{a}_0(\mathbf{X}_{0i}) - n_1^{-1}\hat{a}_0(\mathbf{X}_{1i})\right\},$$

where $\mathbf{X}_{ki}$ is the covariate vector for the ith patient in kth arm, $\hat{a}_k(\mathbf{x}) = \hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}\mathbf{x}$ and $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}_k$ are the least squares estimators of α_k and $\boldsymbol{\beta}_k$ in regression model $E(Y_{ki} \mid \mathbf{X}_{ki}) = \alpha_k + \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{X}_{ki}, k = 0, 1$, respectively. Using the fact that $\sum_{i=1}^{n_k}(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}\mathbf{X}_{ki}) = \hat{\mu}_k$, we have

$$\mu_1^{\dagger} = \pi\hat{\mu}_1 + (1-\pi)(\hat{\alpha}_1 + \hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\bar{\mathbf{X}}_0) \text{ and } \mu_0^{\dagger} = (1-\pi)\hat{\mu}_0 + \pi(\hat{\alpha}_0 + \hat{\boldsymbol{\beta}}_0^{\mathrm{T}}\bar{\mathbf{X}}_1).$$

Since $\hat{\alpha}_k = \hat{\mu}_k - \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}\bar{\mathbf{X}}_k$ and $(\hat{\mu}_k - \mu_k)^2 + (\mu_k^{\dagger} - \mu_k)^2 = o_{a.s.}(n^{-1/2})$,

$$\begin{aligned}\hat{\theta}_{aug} - \hat{\theta} &= -(1-\pi)\dot{g}_2(\hat{\mu}_0, \hat{\mu}_1)\left\{\hat{\mu}_1 - (\hat{\alpha}_1 + \hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\bar{\mathbf{X}}_0)\right\} - \pi\dot{g}_1(\hat{\mu}_0, \hat{\mu}_1)\left\{\hat{\mu}_0 - (\hat{\alpha}_0 + \hat{\boldsymbol{\beta}}_0^{\mathrm{T}}\bar{\mathbf{X}}_1)\right\} + o_{a.s.}(n^{-1/2}) \\ &= -\left\{(1-\pi)\dot{g}_2(\hat{\mu}_0, \hat{\mu}_1)\hat{\boldsymbol{\beta}}_1 - \pi\dot{g}_1(\hat{\mu}_0, \hat{\mu}_1)\hat{\boldsymbol{\beta}}_0\right\}^{\mathrm{T}}\widehat{\Delta}_{\mathbf{X}} + o_{a.s.}(n^{-1/2}).\end{aligned}$$

Now, $\hat{\boldsymbol{\beta}}_k = \hat{\Sigma}_{22k}^{-1}\hat{\Sigma}_{12k}^{\mathrm{T}}$. It follows that

$$\hat{\theta}_{aug} = \hat{\theta} - \left\{(1-\pi)\dot{g}_2(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{121}\hat{\Sigma}_{221}^{-1} - \pi\dot{g}_1(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{120}\hat{\Sigma}_{220}^{-1}\right\}\widehat{\Delta}_{\mathbf{X}} + o_{a.s.}(n^{-1/2}),$$

where $\hat{\Sigma}_{22k}$ is the empirical estimate for $\mathrm{var}(\mathbf{X}|Z = k)$ and $\hat{\Sigma}_{12k}$ is the empirical estimate for $\mathrm{cov}(Y, \mathbf{X}|Z = k)$, $k = 0, 1$. Note that in constructing the bias-adjusted estimator,

$$\begin{aligned}\hat{\Sigma}_{12} &= n^{-1}\left\{\frac{\dot{g}_2(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{121}}{\pi} - \frac{\dot{g}_1(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{120}}{1-\pi}\right\} \text{ and} \\ \hat{\Sigma}_{22} &= n^{-1}\left\{\frac{\hat{\Sigma}_{221}}{\pi} + \frac{\hat{\Sigma}_{220}}{1-\pi}\right\}.\end{aligned}$$

This, coupled with the fact that $\hat{\Sigma}_{221} - \hat{\Sigma}_{220} = o_{a.s.}(1)$, implies that

$$\left\{(1-\pi)\dot{g}_2(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{121}\hat{\Sigma}_{221}^{-1} - \pi\dot{g}_1(\hat{\mu}_0, \hat{\mu}_1)\hat{\Sigma}_{120}\hat{\Sigma}_{220}^{-1}\right\} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1} = o_{a.s.}(1),$$

and

$$\hat{\theta}_{aug} - \hat{\theta}_{adj} = o_{a.s.}(\widehat{\Delta}_{\mathbf{X}} + n^{-1/2}).$$

Therefore

$$\mathrm{pr}\left\{n^{1/2}|\hat{\theta}_{aug} - \hat{\theta}_{adj}| \geq \delta|\widehat{\Delta}_{\mathbf{X}}\right\} = o_{a.s.}(1)$$

as $n \to \infty$ for any positive δ.