数理解析研究所講究録
第1703巻 2010年 53-68

Covariate Adjustment in Randomized Controlled Trial with Binary Outcomes

Masaaki Kuriki, Fumiaki Takahashi, Masahiro Takeuchi
Department of Clinical Medicine(Biostatistics),
Graduate School of Pharmaceutical Sciences, Kitasato University,
5-9-1, Shirokane, Minato-ku, Tokyo, 108-8641, Japan

1 Introduction

The primary outcome is often systematically related to other influences apart from treatment in randomized controlled trials. Randomization tends to produce study groups comparable with respect to known and unknown prognostic factors. Some of these baseline covariates may be related to the primary outcome and may exhibit chance imbalances between the two treatment groups. The primary aim of covariate adjustment is to reduce the bias in the estimate of treatment effect and to improve the precision of that estimate.

For binary outcome case, the most popular covariate adjustment method is logistic regression. Despite this, covariate adjustment with logistic regression leads to a loss of marginal precision in the case of strong outcome-covariate association or strong treatment-covariate association [1]. Logistic regression lead to biased conditional estimates of treatment effect unless the true model is specified (if needed cobariates are omitted) [2]. Marginal treatment effect estimator is not equal to conditional estimator based on covariates.

Koch et al. proposed a nonparametric method for covariate adjustment in randomized controlled trials. This method is based on the assumption that, on average, the two treatment groups are balanced, and will correct for random imbalances between the treatment groups [3]. Tsiatis et al. [4] proposed an adjustment method that follows from application of the theory of semiparametrics by Leon et al. [5]. This approach separates estimation of the treatment difference from the adjustment, which may lessen concerns over bias that could result under regression-based adjustment. Zhang et al. expanded on this idea [6] by developing a broad framework for covariate adjustment in setting with two or more treatments and general outcome summary measures (e.g. log odds ratios) by appealing to the theory of semiparametrics. The estimator gained from this framework yields the greatest efficiency among all estimators in semiparametric class. However, the simulation studies conducted in this article have been only assumed the case of large sample size ($n = 800$). The properties have not been validated in case of small and moderate sample size. Nonparametric covariate adjustment method and semiparametric covariate adjustment method have never been compared in simulation study with binary outcome.

In Section 2.1 and 2.2, we briefly describe nonparametric method [3] and semiparametric method produced [6]. In Section 2.3, we compare performance of these methods in simulation study.

In addition to above topic, we consider stratified randomized controlled trials from semiparametric covariate adjustment method [6]. Stratified randomization is often used to prevent imbalance between treatment groups for known factors that influence prognosis or treatment responsiveness. And stratification may prevent type I error and improve power for small trials (< 400 patients), but only when the stratification factors have a large effect on prognosis [7].

Most multicentre (or multiregional) trials are stratified by centres (or regions) either for practical reasons or because centres are expected to be confounded with other known or unknown prognostic factors. Japanese multicentre trials have many strata and small sample size per stratum. When multicentre trials are stratified by centre and centre effect is not negligible, centre should be adjusted for primary analysis. Similarly, if an alternative feature such as centre (or region) is used as a stratification factor, then this should be adjusted for in the primary analysis [8]. According to the ICH-E9 guideline, multicentre trials is recommended to have large sumple size per stratum in principle (1 group has 10 patients per stratum), and has been based on the use of fixed effect models. However, fixed effect models may be improper in Japanese multicentre trials (when the number of sites is large).

The primary analysis should reflect the restriction on the randomization implied by stratification. When stratified randomization is used, stratification factor is well balanced. However, there is no guarantee that other baseline characteristics will be similar in the treatment groups, especially in case of a small sample within strata. In order to adjust the random imbalance of non-stratified factor, covariate adjustment methods based on stratified randomization are needed. We consider stratified randomized controlled trials from semiparametric covariate adjustment method.

We apply Zhang's semiparametric framework to stratified randomized controlled trials in Section 3.1 and 3.2, and simulation studies demonstrating performance are summarized in Section 3.3.

2 Comparison of Covariate Adjustment Methods in Randomized Controlled Trials with Binary Outcomes

2.1 Koch's nonparametric method for covariate adjustment

Denote the data from a 2-arm randomized trial, as $(Y_i, \boldsymbol{X}_i, Z_i), i = 1, \cdots, n$, independent and identically distributed (i.i.d.) across i, where, for subject i, Y_i is outcome; $\boldsymbol{X}_i = (X_{1i}, \cdots, X_{pi})^T$ is a $(p \times 1)$ vector of all available baseline covariates; and $Z_i = g$ is indicates assignment to treatment group g with known randomization probabilities $P(Z = g) = \pi_g, g = 0, 1$; $Z_i = 0$ for assignment of the ith patient to control group and $Z_i = 1$ for assignment of the ith patient to test treatment group. Randomization ensures that

Z and $\boldsymbol{X}$ are independent. Let the number of subjects randomized to control group and test treatment group be $n_0 = \sum_{i=1}^{n}(1 - Z_i)$, $n_1 = \sum_{i=1}^{n}(Z_i)$ and $n = n_0 + n_1$. Let $\bar{Y}_g$ and $\bar{\boldsymbol{X}}_g$ is denote the sample means of the response variables and the covariables for the patient in the gth group.

Koch et al. proposed a covariate adjustment for comparison of continuous, ordinal and binary responses in a randomized controlled trial [3]. We focus on binary response case. Two approaches are available for estimating covariance structure V. One is a randomization approach while the other is a sampling-based approach that assumes that the study is a sample of a population. Sampling-based approach cannot preserve the probability of the type I error [9]. Thus, we use randomization approach in this study. Koch's method is based on a weighted least-squares procedure to estimate the δ in the linear model

$$E(\boldsymbol{d}) = E\begin{pmatrix} d_Y \\ \boldsymbol{d}_X \end{pmatrix} = E\begin{pmatrix} \bar{Y}_1 - \bar{Y}_0 \\ \bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0 \end{pmatrix} = \begin{pmatrix} \delta \\ \mathbf{0} \end{pmatrix} = \delta\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = \delta\boldsymbol{G},$$

where $\mathbf{0}$ denotes a $(p \times 1)$ vector of 0's. δ is estimated with weights based on a estimate $\boldsymbol{V}$ for the covariance matrix of $\boldsymbol{d}$, which is given by

$$\begin{aligned} \boldsymbol{V} &= \sum_{i=1}^{n} \frac{n}{n_0 n_1 (n-1)} \begin{pmatrix} (Y_i - \bar{Y})^2 & (Y_i - \bar{Y})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \\ (Y_i - \bar{Y})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T & (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \end{pmatrix} \\ &= \begin{pmatrix} V_{YY} & \boldsymbol{V}_{YX}^T \\ \boldsymbol{V}_{YX} & \boldsymbol{V}_{XX} \end{pmatrix}. \end{aligned}$$

The resulting estimator of δ, which is the adjusting difference between the means of the response variable is given by

$$\begin{aligned} \hat{\delta} &= (\boldsymbol{G}^T\boldsymbol{V}^{-1}\boldsymbol{G})^{-1}\boldsymbol{G}\boldsymbol{V}^{-1}\boldsymbol{d} \\ &= d_Y - \boldsymbol{V}_{YX}^T\boldsymbol{V}_{XX}^{-1}\boldsymbol{d}_X \\ &= (\bar{Y}_1 - \bar{Y}_0) - \boldsymbol{V}_{YX}^T\boldsymbol{V}_{XX}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0) \end{aligned}$$

and variance of $\hat{\delta}$ is approximately equal to

$$\mathrm{Var}(\hat{\delta}) = (\boldsymbol{G}^T\boldsymbol{V}^{-1}\boldsymbol{G})^{-1} = V_{YY} - \boldsymbol{V}_{YX}^T\boldsymbol{V}_{XX}^{-1}\boldsymbol{V}_{YX}.$$

A test statistic for comparison of the two treatments with adjustment for the covariates is $T^2_{\mathrm{Koch}} = \hat{\delta}^2/\mathrm{Var}(\hat{\delta})$, it approximately has the chi-square distribution with 1 degree of freedom.

2.2 Zhang's semiparametric method for covariate adjustment

Leon et al. developed the model conceptualizing the situation in terms of counterfactuals or potential outcomes [5]. Ideally, if we could observe outcome Y on each subject under both treatments, we would have complete sample information on treatment effect. Of course, this is usually impossible, but randomization allows us to regard unobserved outcome as 'missing completely at random (MCAR)' [10]. For subjects randomized to

experimental treatment, we observe only their outcome under assigned treatment; the outcome they would have experienced under control is hence 'missing', and vice versa. Covariate adjustment may be viewed as an attempt to use covariates that are correlated with outcome to recover some of the 'lost' information (relative to the 'ideal') due to this 'missingness'. Tsiatis et al. [4] proposed an adjustment method that follows from application of the theory of semiparametrics (e.g. [11, 12]) in order to estimate treatment differences in randomized controlled trials.

Zhang et al. derived the class of estimating functions using the theory of semiparametrics [6]. For binary outcome, one may consider a logistic regression model

$$E(Y|Z;\boldsymbol{\theta}) \;=\; \text{logit}P(Y=1|Z;\boldsymbol{\theta}) = \alpha + \beta Z$$

where $\text{logit}(p) = \log\{p/(1-p)\}$ and β is the log odds ratio for treatment 1 relative to treatment 0. The usual maximum likelihood estimator for β in this model is obtained by solving $\sum_{i=1}^{n} m(Y_i, Z_i; \boldsymbol{\theta}) = 0$, where $\boldsymbol{\theta} = (\alpha, \beta^T)$. The estimating function is written as

$$m(Y_i, Z_i; \boldsymbol{\theta}) \;=\; \begin{pmatrix} 1 \\ Z \end{pmatrix} \left\{ Y - \frac{\exp(\alpha + \beta Z)}{1 + \exp(\alpha + \beta Z)} \right\}.$$

The estimating function for $\boldsymbol{\theta}$ is unbiased and based only on (Y, Z) in model leading to consistent, asymptotically normal estimators. Zhang et al. proposed a fixed unbiased estimating function $m(Y_i, Z_i; \boldsymbol{\theta})(2 \times 1)$ for $\boldsymbol{\theta}$, using all of $(Y, \boldsymbol{X}, Z)$ may be written as

$$m^*(Y_i, \boldsymbol{X}, Z_i; \boldsymbol{\theta}) \;=\; m(Y_i, Z_i; \boldsymbol{\theta}) - \sum_{g=0}^{1} \{I(Z=g) - \pi_g\} q_g^{(2)}(\boldsymbol{X}) \qquad (1)$$

where $q_g^{(2)}(\boldsymbol{X}), g = 0, 1$ are arbitrary 2-dimensional functions of $\boldsymbol{X}$. Because Z and $\boldsymbol{X}$ are independent, the second term in $m^*(Y_i, \boldsymbol{X}_i, Z_i; \boldsymbol{\theta})$ has mean zero; thus, $m^*(Y_i, \boldsymbol{X}_i, Z_i; \boldsymbol{\theta})$ is unbiased estimating function based on $(Y, \boldsymbol{X}_i, Z)$. The adjusted estimator $\hat{\boldsymbol{\theta}}^*$ obtained by solving $\sum_{i=1}^{n} m^*(Y_i, \boldsymbol{X}_i, Z_i; \boldsymbol{\theta}) = 0$ is more efficient than an unadjusted estimator obtained by solving $\sum_{i=1}^{n} m(Y_i, Z_i; \boldsymbol{\theta}) = 0$. When $q_g^{(2)}(\boldsymbol{X}) \equiv \mathbf{0}, g = 0, 1$, estimating function (1) reduces to the original estimating function, which does not take account of auxiliary covariates, and leads to an unadjusted estimator.

An estimator for corresponding to an estimating function in class (1) yields the greatest efficiency gain over $\hat{\boldsymbol{\theta}}$ among all estimators with estimating functions in class (1). The estimator $\hat{\boldsymbol{\theta}}^*$ is consistent and asymptotically normal with asymptotic "sandwich" type covariance matrix

$$\boldsymbol{V}_S^* \;=\; \hat{\boldsymbol{V}_M^*} \hat{\boldsymbol{\Sigma}}_0 \hat{\boldsymbol{V}_M^*} = \begin{bmatrix} \text{Var}(\hat{\alpha}^*) & \text{Cov}(\hat{\alpha}^*, \hat{\beta}^*) \\ \text{Cov}(\hat{\alpha}^*, \hat{\beta}^*) & \text{Var}(\hat{\beta}^*) \end{bmatrix},$$

where $\hat{\boldsymbol{V}_M^*} = [\sum_i \{-\partial/\partial\boldsymbol{\theta}^T m(Y_i, Z_i; \boldsymbol{\theta})\|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^*}\}]^{-1}$, $\hat{\boldsymbol{\Sigma}}_0 = \sum_i \{m^*(Y_i, \boldsymbol{X}, Z_i; \boldsymbol{\theta})^{\otimes 2}\}$, and $\boldsymbol{u}^{\otimes 2} = \boldsymbol{u}\boldsymbol{u}^T$. In general, the optimal estimator in class (1) that has the smallest asymptotic variance is the solution to

$$\sum_{i=1}^{n} \left[m(Y_i, Z_i; \boldsymbol{\theta}) - \sum_{g=0}^{1} \{I(Z_i = g) - \pi_g\} E\{m(Y_i, Z_i; \boldsymbol{\theta}) | \boldsymbol{X}_i, Z_i = g\} \right] \;=\; 0$$

Conditional expectations $E\{m(Y_i, Z_i; \boldsymbol{\theta})|\boldsymbol{X}_i, Z_i = g\}, g = 0, 1$ is replaced by the term $m\{E(Y|\boldsymbol{X}, Z = g), g; \boldsymbol{\theta}\} = (1, g)^T[E(Y|\boldsymbol{X}, Z = g) - \exp(\alpha + \beta g)/\{1 + \exp(\alpha + \beta g)\}]$. We may postulate parametric regression models $E(Y|\boldsymbol{X}, Z = g) = q_g^*(\boldsymbol{X})$ for vector basis function $\boldsymbol{X}$. This representation is called gdirecth implementation strategy in Zhang et al. [6]. The improved estimator $\hat{\boldsymbol{\theta}}^*$ holds consistency and asymptotic normality regardless of whether or not parametric models $q_g^*(\boldsymbol{X})$ are correct specifications of the true $E(Y|\boldsymbol{X}, Z = g)$.

An improved test statistic for the hypothesis that is equal to zero is $T^2_{\text{Zhang}} = \hat{\beta}^{*2}/\text{Var}(\hat{\beta}^*)$, where, it approximately has the chi-square distribution with 1 degree of freedom.

2.3 Simulation Studies I

We report result of simulation based on 5000 Monte Carlo data sets. In simulation I, we focus on estimation of marginal log odds ratio for treatment effect whereas β is defined conditional on X. We considered 2-arms randomized controlled trial with binary response Y. Binary outcome Y_i was generated according to a logistic regression model

$$\text{logit}\{E(Y_i|Z_i)\} = -0.9 + \beta Z_i + \gamma X_i, \tag{2}$$

so that β is the log odds ratio for treatment 1 relative to treatment 0, and γ is covariate effect. For each scenario, treatment indicator Z was generated from Bernoulli with $P(Z = 1) = P(Z = 0) = 0.5$, and covariate X was continuous value generated from standard normal distribution.

In the first set of simulations (scenario A), we considered influence of sample size ($n = 50, 400$) and the strength of association between Y and X ($\gamma = 0.6, 1.2, 1.8$). We estimate the marginal log odds ratio for treatment $Z = 1$ relative to $Z = 0$ in (2) by each method. We compared Monte Carlo bias (MC Bias), Monte Carlo standard deviation (MC SD), the average of estimated standard errors (Ave. SE), the Monte Carlo mean squared error for the unadjusted estimator divided by that for the indicated estimator (RE), empirical size and power and Monte Carlo coverage probability of 95% Wald confidence intervals (CP) computed with the following three methods: 1) the unadjusted estimator based on the data on (Y, Z) (Unadj.), 2) nonparametric method (Koch), 3) semiparametric method (Zhang). To estimate conditional expectations $E\{m(Y_i, Z_i; \theta)|X_i, Z_i = g\}, g = 0, 1$, we develop two types parametric regression models $q_g^*(X)$ based on the data (Y_i, X_i) for i in group g: 3-A) Logit type, $E(Y_i|X_i, Z_i = g) = q_g^*(X) = \text{logit}(\zeta_0 + X_i\zeta_g)$ (Zhang-A), 3-B) Linear type, $E(Y_i|X_i, Z_i = g) = q_g^*(X) = \zeta_0 + X_i\zeta_g$ (Zhang-B)

Table 1 shows results of precision comparison for the first set ($\beta = 0.6$). As all estimators showed negligible bias, bias is not reported. For each case, both covariate adjustment methods (Koch and Zhang-A, B) yield considerable efficiency and precision gain over the unadjusted estimator. Comparison of the average of the standard error and Monte Carlo standard deviation shows that sandwich type variance estimator is underestimated in the case of small sample size ($n = 50$). And Zhang's Wald confidence interval estimated with Zhang's method does not achieve the nominal 95% level in our simulation. Koch's method has more precise estimator than Zhang's method in small sample case ($n = 50$). In moderate-size and moderate-strong association ($n = 400, \gamma = 1.2, 1.8$), Zhang's method performs more precise estimation than Koch's estimator, this

results confirm a theory of semiparametric efficiency. Table 2 shows that Zhang's Wald type tests yield greater power than Koch's method. But Zhang's method does not achieve the 5% nominal level in small sample size case ($n = 50$). In contrast, Koch's test achieves the 5% nominal level in all situations.

In the second set of simulation (scenario B), we considered a logistic regression model that there are some interactions between treatment Z and covariate X.

$$\text{logit}\{E(Y_i|Z_i)\} = -0.9 + \beta Z_i + \gamma X_i + \tau(Z_i \times X_i),$$

where τ is the interaction factor. Table 3 shows the performance of the semiparametric method in such circumstance.

In moderate-size ($n = 400$), Zhang's methods more precise estimation than Koch's estimator, depending on the strength of the interaction factor. The optimal $E(Y|X, Z = g)$ are the true regression relationships of Y on X for each treatment 'separately'. With semiparametric covariate adjustment method, the regression relationship of Y on X is constructed for each treatment separately. Thus, semiparametric method give good performance under the condition such as interactions between treatment Z and covariate X.

3 Extensions for stratified studies

Stratified analysis offer three distinct advantages over non-stratified analysis in clinical trial: 1) it enables us to eliminate the variation between strata, 2) it provides comparison of broad patient population without loss of precision, and 3) it provides a broad patient population to support generalizability of findings. In order to extend Zhang's framework to stratified analysis, we produce two strategies due to sample size per stratum. The first is combining strata based on weighted average for large sample size per stratum. The second is application of the framework to conditional logistic regression for small sample size per stratum.

3.1 Weighted average strategy (Large sample size per stratum)

In the case of clinical trials with at least moderately large sample per stratum (e.g. $n_h \geq 50$), $\hat{\beta}_h^*$ and V_{Sh}^* is estimated by using $\boldsymbol{Y}_h, \boldsymbol{X}_h$ and bZ_h via the counterparts of (1) from the hth stratum. Let w_h be a standardized weight for the hth stratum such that $\sum_{h=1}^{q} w_h = 1$. Analysis for combined strata is then based on $\hat{\beta}^* = \sum_{h=1}^{q} w_h \hat{\beta}_h^*$ and $V_S = \sum_{h=1}^{q} w_h^2 V_{Sh}^*$, where

$$w_h = \frac{1/V_{Sh}^*}{\sum_{h'=1}^{q} 1/V_{Sh'}^*}$$

This estimator based on w_h is the best (i.e. minimum variance) weighted average to estimate for a common mean [13].

3.2 Application to conditional logistic regression (Small sample size per stratum)

Above weighted average strategy is improper in case of small sample size per stratum. Because standard asymptotic properties of covariate adjustment estimators from every stratum do not hold, when the number of strata is large and the data are sparse. And the number of parameters grows at the same rate as the number of strata. Those situations commonly occur in the multicentre trials. When multicentre trials are stratified by centre, some centres have few patients (e.g. Japanese multicentre trials).

In order to extend semiparametric model framework to the stratified randomized controlled trial with small sample size per stratum, we apply the framework to conditional logistic regression model. Hauck recommended using the conditional maximum likelihood estimator (i.e. conditional logistic regression) or the Mantel-Haenszel estimator of common odds ratio [14]. These estimators have good asymptotic properties for both asymptotic cases: 1) the number of strata is fixed and sample size within each stratum becomes large, 2) the stratum sizes are fixed, but the number of strata becomes large. We consider a conditional logistic model

$$\text{logit}\{E(Y|Z,h;\beta)\} = \alpha_h + \beta Z, \tag{3}$$

where the nuisance parameters α_h are the stratum-specific intercepts. Approximate likelihood [15] for this model is written as

$$L(\beta) = \prod_{h=1}^{q} L(\beta; \boldsymbol{Z}, \boldsymbol{Y}, h) = \prod_{h=1}^{q} \frac{\prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_{hi} Y)}{\sum_c \prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_{hi})},$$

$$\log L(\beta) = \sum_{h=1}^{q} \log L(\beta; \boldsymbol{Z}, \boldsymbol{Y}, h) = \sum_{h=1}^{q} \left[\sum \beta Z_{hi} Y_{hi} - \log \left\{ \sum_c \prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_{hi}) \right\} \right]$$

where $m_h = \sum_{i=1}^{n_h} Y_{hi}$, the summation is over all combination $c = (c_1, \cdots, c_{m_h})$ of m_h "cases" chosen from the n_h individuals in the stratum. About computation of the conditional likelihood function [16]. The score function and information function are given by

$$U(\beta) = \frac{\partial}{\partial \beta} \log L(\beta) = \sum_{h=1}^{q} \left[\sum_{i=1}^{n_h} \beta Z_{hi} Y_{hi} - \log \left\{ \sum_c \prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_{hi}) \right\} \right]$$

and

$$I_0(\beta) = -\frac{\partial^2}{\partial \beta^2} \log L(\beta) = \sum_{h=1}^{q} D_h$$

Thus, the estimating function may be written as

$$m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) = \sum_{i=1}^{n_h} Z_{hi} Y_h i - \frac{\sum_c \left(\sum_{i=c_1}^{c_{m_h}} Z_h i \right) \prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_h i)}{\sum_c \prod_{i=c_1}^{c_{m_h}} \exp(\beta Z_h i)}$$

maximum partial likelihood estimator (MPLE) for in this model is obtained by solving $\sum_{h=1}^{q} m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) = 0$. The adjusted estimator obtained by solving $\sum_{h=1}^{q} m(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \beta) = 0$ is more efficient than an unadjusted estimator obtained by solving $\sum_{h=1}^{q} m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) = 0$. We used a Newton-Raphson algorithm to solve the estimating equation. Stratified Zhang's unbiased estimating function is written as

$$m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \beta) = m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) - \sum_{g=0}^{1} [\{I(Z = g) - \pi_{hg}\} \times a_g(\boldsymbol{X}_{hi})]$$

where $a_g(\boldsymbol{X}_{hi}), g = 0, 1$ are arbitrary functions of $\boldsymbol{X}$. The optimal estimating function is given by

$$\begin{aligned} m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \beta) &= m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) - \sum_{g=0}^{1} \Big[\{I(Z = g) - \pi_{hg}\} \\ &\quad \times E\{m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) | \boldsymbol{X}_h, \boldsymbol{Z}_h = g\}\Big] \qquad (4) \\ &= m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta) - \frac{n_{h1} n_{h0}}{n_h} (\bar{q}^*_{1(h1)} - \bar{q}^*_{1(h0)}) \end{aligned}$$

where $\bar{q}^*_{1(hg)} = n_{hg}^{-1} \sum_{n_h}^{-1} I(Z_{hi} = g) q_1^*(\boldsymbol{X}_h i), q_1^*(\boldsymbol{X}_h i) = E(Y | \boldsymbol{X}, \boldsymbol{Z} = g, h; \zeta_1)$. In general, $\hat{\beta}_{\text{opt}}$ that has smallest asymptotic variance in class (4) is the solution to $\sum_{h=1}^{q} m(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \beta) = 0$. We develop a parametric regression model based on the data $(Y_i, \boldsymbol{X}_i)$ for i in treatment group $(Z = 1)$

$$q_1^*(\boldsymbol{X}_{hi}) = E\{Y_i | \boldsymbol{X}_i, Z_i = 1, h; \zeta_1) = \text{logit}(\alpha + \boldsymbol{X}_i^T \zeta_1 + u_h)$$

where $\zeta_g = (\zeta_{g1}, \cdots, \zeta_{gp})$, u_h are the stratum-specific random effects. We obtain the "sandwich" type estimator of variance as a function of β,

$$\begin{aligned} V_S^* &\approx \{I_0(\hat{\beta}^*)\}^{-1} \text{Cov}\{U(\hat{\beta}^*)\} \{I_0(\hat{\beta}^*)\}^{-1} \\ &= \{\sum_{h=1}^{q} \hat{D}_h^*\}^{-1} \left[\sum_{h=1}^{q} \{m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \hat{\beta}^*)\}\right] \{\sum_{h=1}^{q} \hat{D}_h^*\}^{-1} \\ &= V_M^* \left[\sum_{h=1}^{q} \{m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \hat{\beta}^*)\}\right] V_M^* \qquad (5) \end{aligned}$$

The classical variance estimator (model-based variance estimator) for conditional logistic regression is $V_M^* = \{\sum_{h=1}^{q} \hat{D}_h^*\}^{-1}$, which can be related to equation (5) by substituting $\sum_{h=1}^{q} \hat{D}_{h=1}^*$ for $\sum_{h=1}^{q} \{m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \hat{\beta}^*)\}^2$. However, the empirical estimator $\sum_{h=1}^{q} \{m^*(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, h; \hat{\beta}^*)\}^2$ is a fairly good estimator of the $\text{Cov}\{U(\hat{\beta}^*)\}$ as long as $q \gg 1$ and the sandwich type estimator underestimates the variance of an estimated stratum-level fixed effect when there are a small number of strata [17]. Even when $q \gg 1$, the middle of the sandwich estimator may be quite unstable and have substantial bias. Wald type sandwich tests tend to be liberal [18, 19, 20, 21, 22, 23, 24]. In order to correct for the sandwich type estimator, we applied the adjustment methods proposed by Fay

et al. [25], which is applied to conditional logistic regression. This method was proposed as two type of adjustments; 1) using Taylor series approximations to adjust for the bias of the sandwich estimator of variance, 2) using a t distribution instead of a chi-square distribution to calculate significance.

3.3 Simulation Studies II

We report result of simulation based on 5000 Monte Carlo data sets. We focus on estimation of log odds ratio for treatment effect conditional on strata h. We considered 2-arms stratified randomized controlled trial with binary response $Y(n = 400)$. $h(= 1, \cdots, q)$ is strata indicator. All strata have the equivalent number of subjects $400/q$. Binary outcome Y_{hi} was generated according to a mixed effects logistic regression model

$$\text{logit}\{E(Y_{hi}|Z_{hi}\} \quad = \quad -0.9 + \beta Z_{hi} + \gamma X_{hi} + b_h$$

for $\beta = 0, 0.6$ and $\gamma = 0.9, 1.5, 2.1$. The b_h is a random strata effect. Random strata effects were generated from a normal distribution with mean 0 and variance 1. We simulated for two scenarios of stratified studies: A) large sample size per stratum ($q = 2, 4$), B) small sample size per stratum ($q = 10, 40$). For each scenario, we applied "permuted block randomization (block of size 4)" to avoid serious treatment imbalance within strata.

We estimate the common log odds ratio for treatment $Z = 1$ relative to $Z = 0$ in (3) by each method. We compared Monte Carlo bias, average length of 95% Wald confidence interval (L), empirical size and power, Monte Carlo coverage probability of 95% Wald confidence intervals (CP), and the Monte Carlo mean squared error for the unadjusted estimator divided by that for the indicated estimator (RE) computed with the following four methods in two scenarios: For scenario C, we compared three methods: 1) Mantel-Haenszel Method (MH), 2) Conditional logistic regression unadjusted for covariate X (CL), 3) Stratified semiparametric method with weighted mean (Zhang-W). In order to represent conditional expectations $E\{m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta)|\boldsymbol{X}_h, \boldsymbol{Z}_h = g\}, g = 0, 1$, we develop a logistic model for each stratum: $E(Y_i, X_i, Z_i = g; h) = \text{logit}(\zeta_{0hg} + X_i^T \zeta_{hg})$ using SAS 9.1 PROC LOGISTIC.

For scenario D, we compared four methods: 1) Mantel-Haenszel Method (MH), 2) Conditional logistic regression unadjusted for covariate X (CL), 3) Stratified semiparametric method with bias-corrected variance estimator proposed by Fay et al. (Zhang-F). In order to represent conditional expectations $E\{m(\boldsymbol{Y}, \boldsymbol{Z}, h; \beta)|\boldsymbol{X}_h, \boldsymbol{Z}_h = g, h\}, g = 0, 1$, we develop a logistic model with random strata effects u_{gh}: $E(Y_i|X_i, Z_i = g; h) = \text{logit}(\zeta_{0hg} + X_i^T \zeta_{hg} + u_{gh})$ using SAS 9.1 PROC NLMIXED.

Table 4 shows results for scenario C: $q = 2, 4$ strata studies. "NC (Non convergence, %)" is the proportion of the cases that logistic regression could not computed in any stratum. These cases were excluded from the summary. As all estimators showed negligible bias, bias is not reported. For both cases, proposed estimator has considerable efficiency which gains over the estimator unadjusted covariate X (Mantel-Haenszel method and conditional logistic regression) and its 95% confidence interval covered the true value at 95% nominal level. Table 5 shows that proposed method can preserve the probability of the type I error. Additionally, proposed method performs more precise estimation in fewer strata.

Table 6 shows results for scenario D: highly stratified studies with small sample sizes per stratum ($q = 10, 40$ strata). All estimators are unbiased. For RE, the proposed method has more precise point estimator than Mantel-Haenszel method and conditional logistic regression. In respect of interval estimate, the proposed estimator has narrower confidence interval than Mantel-Haenszel method and conditional logistic regression when the study has a number of strata and covariate effect is not weak. Table 7 shows that the proposed Wald test achieves the nominal level with Fay's bias correction in all situations. In contrast, the nominal level does not preserved with Morel's bias correction in case of small strata ($q = 10$). When the study has a number of strata and covariate effect is not weak, the proposed test increases in power over the competitors.

4 Discussion

The unconditional treatment effect is overwhelmingly the focus of the primary analysis in most randomized trial. Treatment effect estimator (log odds ratio) adjusted for covariate X (conditional on X) by logistic regression model is not equal to marginal estimator adjusted by Koch's nonparametric method or Zhangfs semiparametric method [14]. Thus we did not compare logistic regression with these methods. In the simulation study I, both Koch's nonparametric method and Zhang's semiparametric method provided more precise estimator than unadjusted estimator in every case. Especially, we recommend that using Koch's nonparametric method is better than Zhang's semiparametric method, because 1) Koch's nonparametric method has fewer assumptions (only randomization and no arbitrary modeling step), 2) in the case of small sample size ($n = 50$), Koch's nonparametric method is superior in terms of MSE, and 3) nonparametric test achieves the 5% nominal level in all situations. Tsiatis et al. compared these methods with continuous outcome, and semiparametric method is superior to nonparametric method in terms of both MSE and Ave. SE in moderate-size trial (n=400) [4].

We have applied semiparametric model framework to stratified randomized controlled trials by using two strategies. Proposed method with weighted mean was shown to be of benefit in large sample size per stratum situation (e.g. stratified by disease stage). When each stratum has the large sample size, it is ensured that semiparametric method proposed by Zhang et al. is proper in each stratum. Properness of the proposed method with weighted mean depends on the sample size per stratum. That kind of stratified randomized controlled trial is recommended, which keeps the number of strata small [7].

In the case of multicentre trials, which have many strata and small sample size per stratum, above proposed method is not available. Second proposed method, which is applied to conditional logistic model, is useful method when centre effect is not negligible (e.g. postoperative chemotherapy trial) and the trial is highly stratified. Properness of this method depends on the number of strata. Wald test based on sandwich type variance estimator may have greater than nominal size if the number of strata is small. In order to compute standard errors, we use the bias correction proposed by Fay et al. The corrected sandwich estimator has conservative confidence intervals when the trial has few strata. However, proposed method does not the smallest length of the confidence interval when the number of strata is small and covariate effect is not strong. This problem is topic for

further research.

We need to demonstrate the proposed method to competing methods by application to clinical trial data. We considered the case of only one covariate effect in simulation studies. Using the coefficient of determination R^2 to measure the strength of covariate effect, R^2 is large value when outcome is related to a number of covariates. So we expected that proposed method is more effective in real clinical trial data analysis than simulation studies.

References

[1] Robinson, L. D. and Jewell, N. P. (1991) Some Surprising Result About Covariate Adjustment in Logistic Regression Models. International Statistical Review 58:227-240.

[2] Gail, M. H., Wieand, S. and Piantadosi, S. (1984) Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 71: 431-444.

[3] Koch, G. G., Tangen, C. M., Jung, J. W. and Amara, I. A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine 17:1863-1892.

[4] Tsiatis, A. A., Davidian, M. Zhang, M. and Lu, X. (2007) Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. Statistics in Medicine 27:4658-4677.

[5] Leon, S., Tsiatis, A. A. and Davidian, M. (2003) Semiparametric Estimation of Treatment Effect in a Pretest-Posttest Study. Biometrics 59:1046-1055.

[6] Zhang, M., Tsiatis, A. A. and Davidian, M. (2008) Improving Efficiency of Inference in Randomized Clinical Trials Using Auxiliary Covariates. Biometrics 64:707-715.

[7] Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. and Horwitz, R. I. (1999) Stratified Randomization for Clinical Trials. Journal of Clinical Epidemiology 52:19-26.

[8] Committee for Proprietary Medical Products (2003) Point to consider on adjustment for baseline covariate. CPMP/EWP/2863/99

[9] Lesaffre, E. and Senn, S. (2003) A note on non-parametric ANCOVA for covariate adjustment in randomized clinical trials. Statistics in Medicine 22:3583-3596.

[10] Rubin, D. B. (1976) Inference and missing data. Biometrika 63: 581-592.

[11] van der Laan, M. J. and Robins, J. M. (2003) Unified Methods for Censored Longitudinal Data and Causality. New York: Springer.

[12] Tsiatis, A. A. (2006) Semiparametric Theory and Missing Data. New York: Springer.

[13] Lehmann, E. L. (1999) Elements of large-sample theory. New York: Springer.

[14] Hauck, W. W. (1989) Odds ratio inference from stratified samples. Communications in Statistics Theory and Methods 18:767-800.

[15] Cox, D. R. (1972) Regression Models and Life-Tables. Journal of the Royal Statistical Society 34: 187-220.

[16] Gail, M. H., Lubin, J. H. and Rubinstein L. V. (1981) Likelihood calculations for matched case-control studies and survival studies with tied death times. Biometrika 68: 703-707.

[17] Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994) Analysis of longitudinal data. New York: Oxford.

[18] Lin, D. Y. and Wei, L. J. (1989) The robust inference for the Cox proportional hazards model. Journal of the American Statistical Association 84: 1074-1078.

[19] Emrich, A. H-T. and Piedmonte, M. R. (1992) On some small sample properties of generalized estimating equation estimates for multivariate dichotomous outcomes. Journal of Statistical Computation and Simulation 41: 19-29.

[20] Gunsolley, J. C., Getchell, C. and Chinchilli, V. M. (1995) Small sample characteristics of generalized estimating equations. Communications in Statistics: Simulation and Computation 24: 869-878.

[21] Fay, M. P., Graubard, B. I., Freedman, L. S. and Midthune, D. N. (1998) Conditional logistic regression with sandwich estimators: application to a meta-analysis. Biometrics 54: 195-208.

[22] Mancl, L. A. and DeRouen, T. A. (2001) A covariance estimator for GEE with improved small sample properties. Biometrics 55: 574-579.

[23] Pan, W. and Wall, M. M. (2002) Small-sample adjustment in using the sandwich variance estimator in generalized estimating equations. Statistics in Medicine 21: 1429-1441.

[24] Morel, J. G., Bokossa, M. C. and Neerchal, N. K. (2003) Small Sample Correction for the Variance of GEE Estimators. Biometrical Journal 45: 395-409.

[25] Fay, M. P. and Graubard, B. I. (2001) Small-Sample Adjustments for Wald-Type Tests Using Sandwich Estimators. Biometrics, 57:1198-1206.

Table 1: Simulation results for estimation of the log odds ratio for treatment $Z = 1$ relative to $Z = 0$ based on 5000 Monte Carlo data sets. (scenario A)

	$n = 50$				$n = 400$			
Estimator	MC SD	Ave. SE	RE	CP	MC SD	Ave. SE	RE	CP
				$\beta = 0.6, \gamma = 0.6$				
Unadj.	0.637	0.621	1	0.956	0.21	0.211	1	0.952
Koch	0.609	0.592	1.1	0.956	0.203	0.203	1.07	0.951
Zhang-A	0.622	0.591	1.05	0.95	0.203	0.203	1.07	0.952
Zhang-B	0.625	0.593	1.04	0.951	0.203	0.203	1.07	0.951
				$\beta = 0.6, \gamma = 1.2$				
Unadj.	0.643	0.656	1	0.957	0.208	0.208	1	0.953
Koch	0.547	0.538	1.38	0.956	0.185	0.185	1.27	0.951
Zhang-A	0.551	0.531	1.36	0.948	0.184	0.184	1.28	0.953
Zhang-B	0.56	0.538	1.32	0.95	0.185	0.185	1.27	0.951
				$\beta = 0.6, \gamma = 1.8$				
Unadj.	0.626	0.602	1	0.953	0.206	0.206	1	0.950
Koch	0.507	0.49	1.52	0.951	0.168	0.169	1.51	0.951
Zhang-A	0.511	0.474	1.5	0.939	0.166	0.165	1.54	0.947
Zhang-B	0.521	0.489	1.44	0.944	0.169	0.169	1.5	0.949

Table 2: Proportion of 5000 Monte Carlo data sets for which the null hypothesis $\beta = 0$ is rejected in favor of the alternative $\beta \neq 0$ using the test statistic based on each estimator and level of significance 0.05. (scenario A)

		$\beta = 0$				$\beta = 0.6$			
n	γ	Unadj.	Koch	Zhang-A	Zhang-B	Unadj.	Koch	Zhang-A	Zhang-B
50	0.6	0.041	0.047	0.047	0.045	0.148	0.159	0.166	0.165
	1.2	0.043	0.046	0.055	0.051	0.114	0.137	0.145	0.148
	1.8	0.05	0.048	0.059	0.056	0.101	0.13	0.146	0.142
400	0.6	0.049	0.046	0.046	0.046	0.757	0.784	0.787	0.784
	1.2	0.048	0.053	0.055	0.053	0.63	0.726	0.733	0.728
	1.8	0.048	0.044	0.047	0.045	0.463	0.627	0.645	0.629

Table 3: Results for the second simulation scenario data using 5000 Monte Carlo data sets. (scenario B)

	$n = 400$			
Estimator	MC SD	Ave. SE	RE	CP
	$\beta = 0.6, \gamma = 0.9, \tau = 0$			
Unadj.	0.209	0.209	1	0.952
Koch	0.194	0.194	1.16	0.949
Zhang-A	0.194	0.194	1.15	0.949
Zhang-B	0.195	0.194	1.15	0.949
	$\beta = 0.6, \gamma = 0.9, \tau = 0.3$			
Unadj.	0.211	0.209	1	0.952
Koch	0.192	0.191	1.21	0.949
Zhang-A	0.192	0.19	1.21	0.950
Zhang-B	0.192	0.191	1.2	0.949
	$\beta = 0.6, \gamma = 0.9, \tau = 0.6$			
Unadj.	0.212	0.209	1	0.950
Koch	0.19	0.187	1.24	0.946
Zhang-A	0.19	0.186	1.25	0.947
Zhang-B	0.191	0.187	1.23	0.946
	$\beta = 0.6, \gamma = 0.9, \tau = 0.9$			
Unadj.	0.209	0.209	1	0.954
Koch	0.186	0.185	1.27	0.949
Zhang-A	0.185	0.184	1.28	0.947
Zhang-B	0.186	0.185	1.26	0.947

Table 4: Simulation results for estimation of the log odds ratio for treatment $Z = 1$ relative to $Z = 0$ based on 5000 Monte Carlo data sets. (scenario C; $n = 400$) NC (Non convergence): Proportion of the case in which Weight method could not compute the common log odds estimator (all observation has same response in any group and stratum).

	$q = 2$			$q = 4$		
Estimator	L	RE	CP	L	RE	CP
	$\beta = 0.6, \gamma = 0.9$					
MH	0.886	1.00	0.952	0.880	1.00	0.952
CL	0.885	1.01	0.952	0.878	1.01	0.952
Wt	0.826	1.15	0.951	0.821	1.14	0.952
	NC : 0.1%			NC : 1.46%		
	$\beta = 0.6, \gamma = 1.5$					
MH	0.857	1.00	0.954	0.854	1.00	0.953
CL	0.856	1.01	0.954	0.853	1.01	0.952
Wt	0.728	1.37	0.951	0.724	1.35	0.950
	NC : 0.04%			NC : 0.46%		
	$\beta = 0.6, \gamma = 2.1$					
MH	0.835	1.00	0.956	0.834	1.00	0.956
CL	0.834	1.01	0.956	0.833	1.01	0.956
Wt	0.645	1.65	0.945	0.640	1.61	0.948
	NC = 0.02%			NC = 0.14%		

Table 5: Proportion of 5000 Monte Carlo data sets for which the null hypothesis $\beta = 0$ is rejected in favor of the alternative $\beta \neq 0$ using the test statistic based on each estimator and level of significance 0.05. (scenario C)

		$\beta = 0$			$\beta = 0.6$		
q	γ	MH	CL	Wt	MH	CL	Wt
2	0.9	0.050	0.050	0.050	0.650	0.649	0.709
	1.5	0.047	0.046	0.049	0.515	0.514	0.653
	2.1	0.045	0.045	0.048	0.391	0.389	0.591
4	0.9	0.049	0.049	0.050	0.645	0.644	0.694
	1.5	0.048	0.047	0.053	0.520	0.518	0.651
	2.1	0.047	0.046	0.053	0.392	0.390	0.585

Table 6: Simulation results for estimation of the conditional log odds ratio for treatment $Z = 1$ relative to $Z = 0$ based on 5000 Monte Carlo data sets. (scenario D; $n = 400$)

	$q = 10$			$q = 40$		
Estimator	RE	L	CP	RE	L	CP
			$\beta = 0.6, \gamma = 0.9$			
MH	1.00	0.878	0.953	1.00	0.887	0.948
CL	1.03	0.877	0.9522	1.01	0.884	0.950
Zhang-F	1.17	0.928	0.9526	1.12	0.848	0.948
			$\beta = 0.6, \gamma = 1.5$			
MH	1.00	0.852	0.950	1.00	0.860	0.953
CL	1.01	0.851	0.950	1.01	0.858	0.954
Zhang-F	1.37	0.818	0.950	1.36	0.756	0.951
			$\beta = 0.6, \gamma = 2.1$			
MH	1.00	0.833	0.947	1.00	0.841	0.955
CL	1.01	0.832	0.947	0.93	0.839	0.954
Zhang-F	1.65	0.733	0.949	1.45	0.675	0.950

Table 7: Proportion of 5000 Monte Carlo data sets for which the null hypothesis $\beta = 0$ is rejected in favor of the alternative $\beta \neq 0$ using the test statistic based on each estimator and level of significance 0.05. (scenario D; $n = 400$)

		$\beta = 0$			$\beta = 0.6$		
q	γ	MH	CL	Zhang-F	MH	CL	Zhang-F
10	0.9	0.048	0.048	0.051	0.649	0.648	0.596
	1.5	0.0472	0.047	0.0486	0.514	0.513	0.544
	2.1	0.047	0.047	0.047	0.398	0.398	0.482
40	0.9	0.056	0.057	0.051	0.650	0.649	0.690
	1.5	0.045	0.045	0.050	0.503	0.503	0.612
	2.1	0.041	0.042	0.048	0.390	0.390	0.551