Research Article

Statistics in Medicine

Received 25 April 2014, Accepted 15 January 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6447

The impact of covariate adjustment at randomization and analysis for binary outcomes: understanding differences between superiority and noninferiority trials

Katherine Nicholas,[a*†] Sharon D. Yeatts,[a] Wenle Zhao,[a] Jody Ciolino,[b] Keith Borg[c] and Valerie Durkalski[a]

The question of when to adjust for important prognostic covariates often arises in the design of clinical trials, and there remain various opinions on whether to adjust during both randomization and analysis, at randomization alone, or at analysis alone. Furthermore, little is known about the impact of covariate adjustment in the context of noninferiority (NI) designs. The current simulation-based research explores this issue in the NI setting, as compared with the typical superiority setting, by assessing the differential impact on power, type I error, and bias in the treatment estimate as well as its standard error, in the context of logistic regression under both simple and covariate adjusted permuted block randomization algorithms.

In both the superiority and NI settings, failure to adjust for covariates that influence outcome in the analysis phase, regardless of prior adjustment at randomization, results in treatment estimates that are biased toward zero, with standard errors that are deflated. However, as no treatment difference is approached under the null hypothesis in superiority and under the alternative in NI, this results in decreased power and nominal or conservative (deflated) type I error in the context of superiority but inflated power and type I error under NI. Results from the simulation study suggest that, regardless of the use of the covariate in randomization, it is appropriate to adjust for important prognostic covariates in analysis, as this yields nearly unbiased estimates of treatment as well as nominal type I error. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: noninferiority; clinical trials; randomization; covariates

1. Introduction

The noninferiority (NI) trial design is growing progressively more popular as the need for comparable therapies with secondary advantages increases. The challenges that arise in the design and analysis of such a trial have been discussed in the Food and Drug Administration Guidance on NI trials as well as methodological research on proper specification of hypotheses, the choice of active control, the determination of the NI margin, and the appropriate analysis method [1–7]. However, critical gaps in the literature remain regarding key design issues, specifically the impact of covariate adjustment at both randomization and at the analysis phase.

In a PubMed search of phase-III clinical trials published in the last 10 years containing the word 'noninferiority', 44.8% of 58 trials utilized simple randomization and unadjusted primary analyses. Covariate adjusted randomization and appropriate analysis of covariance or stratification was used in 20.7% of trials. Covariate adjustment at randomization but not at analysis was done in 25.7%, and adjustment at

[a]*Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, U.S.A*
[b]*Department of Preventive Medicine, Northwestern University, Evanston, IL, U.S.A*
[c]*Department of Pediatrics/Emergency Medicine, Medical University of South Carolina, Charleston, SC, U.S.A*
**Correspondence to: Katherine Nicholas, Department of Public Health Science, Medical University of South Carolina, 135 Cannon Street, Suite 305R, Charleston, SC 29425, U.S.A.*
[†]*E-mail: nicholk@musc.edu*

analysis but not at randomization was done in 8.6% of the trials. This review of current practice shows that there is still considerable variability in the way that clinical trialists use covariate information, a finding that is not surprising considering the lack of attention it has received in the NI literature.

This manuscript expands the research on covariate adjustment in the NI setting by examining the impact of adjustment at randomization as well as at analysis using logistic regression models. A simulation study is conducted to examine the operating characteristics in both superiority and NI settings. Section 2 reviews the existing statistical and clinical literature on covariate adjustment. Sections 3 and 4 present the simulation methods and results, and Section 5 discusses the differential impact of covariate adjustment in these two settings.

2. Existing literature

Little work has been published to date to examine the impact of covariate adjustment at both randomization and analysis in the context of NI trials. Garrett [4] explains that, due to the reversal of hypotheses from the superiority setting, there is also a reversal of the impact of errors. Thus, he cautions readers that the power and type I error is inflated when important prognostic factors are ignored in the NI setting. [4] However, this work does not take into account the potential impact of the randomization scheme.

It has been shown in a superiority setting that failure to adjust for important prognostic covariates at either randomization or analysis leads to biased estimates of the treatment effect, the direction and magnitude of which is dependent on the strength of association between the covariate and outcome, as well as the level of covariate imbalance across treatment arms [8–16]. Gail *et al.* [13] illustrate the nature of this bias via Taylor series approximation in a nonlinear setting. They show that in a linear setting, the estimated treatment effect in an adjusted versus in an unadjusted model are equivalent. In the nonlinear (i.e. the logistic) setting, they quantify the discrepancy between the adjusted treatment effect estimate and the unadjusted treatment effect estimate and demonstrate that this discrepancy is nonzero (i.e. the unadjusted effect will always be biased) except when there is no treatment effect, when there is no association between covariate and outcome, or when the variance of the covariate is zero. Furthermore, Gail *et al.* [14] show that the bias tends to be such that the unadjusted effect will underestimate the adjusted effect if the treatment effect is positive, resulting in a decrease in power, and this effect holds even when covariates are perfectly balanced. As Robinson and Jewell [15] point out, this underestimation of the treatment effect when we fail to adjust for covariates outweighs any benefit in standard error [13–18].

The impact of covariate adjustment at randomization has been evaluated in the context of superiority. Kahan and Morris [8] illustrate that, for continuous, binary, and time-to-event outcomes, stratified randomization creates correlation between treatment arms. Thus, unadjusted analyses in this context result in decreased power and type I error, as well as inflated standard errors of the treatment effect [8].

In light of the statistical evidence in favor of adjusted analyses, it is clear that important prognostic covariates that are included in the randomization scheme should also be included in the final analysis in the form of a properly specified analysis of covariance [2, 10, 19], and in fact, this results in unbiased estimates of treatment effect, as well as nominal power and type I error. However, the full impact of covariate adjustment in the NI setting remains to be demonstrated and is a topic worthy of further investigation.

3. Simulation studies

Parameters are specified according to both the hypothesis of interest (null or alternative) and the scenario of interest (superiority or NI). Four simulation studies were designed to perform simple randomization and covariate adjusted permuted block randomization in the context of superiority and NI designs. All simulations conducted both unadjusted and adjusted analyses based on the following models:

$$\text{Unadjusted}: \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 Trt \tag{1}$$

$$\text{Adjusted}: \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 Trt + \beta_2 \text{Covariate} \tag{2}$$

The probability of success in the control group (π_C) was set at 80%. This value was chosen based on an NI setting, where the active control may have a high probability of success; however, the experimental treatment offers other advantages such as lower side effects.

3.1. Simulation parameters

The probability of success in the treatment group (π_T), the sample size, and the predetermined margin were differentially specified depending on the statistical hypothesis to be tested (superiority or NI), but the simulation strategy remained the same. For the superiority setting, the probability of success in the treatment group was set at 90% in order to mimic a trial with an expected absolute difference in treatment of 10%. The null hypothesis is $H_0 : \pi_C = \pi_T = 0.80$, and the alternative is $\pi_C = 0.80$ and $\pi_T = 0.90$. Thus, β_1 is estimated to be $\ln(1) = 0$ under the null hypothesis and $\ln(2.25) = 0.811$ under the alternative. This information was then used to calculate the total sample size of $N = 392$ subjects in order to ensure power of 80% when there was no effect of the covariate ($\beta_2 = 0$).

The value of β_2 was allowed to vary from −3.0 to 3.0 in increments of 0.5 in order to evaluate the effect of strength of covariate on operating characteristics, and Equations (1) and (2) earlier were used to derive the intercept, β_0, for each scenario. Although it is not expected that any covariate would have a coefficient as high as 3.0, this range was included in the simulation to examine the theoretical setting.

For the case of NI, we set the NI margin at 0.10, resulting in the following hypotheses:

$$H_0 : \ \pi_C - \pi_T \geqslant 0.10 \tag{3}$$

$$H_1 : \ \pi_C - \pi_T < 0.10 \quad , \tag{4}$$

where $\pi_C = 0.80$ and $\pi_T = 0.70$ under the null and $\pi_C = \pi_T = 0.80$ under the alternative. This yielded a β_1 under the null of −0.539 and under the alternative a value of 0. An odds ratio of 0.583 was derived from the expected probability of success in each of the two groups, which served as the cutoff for claiming NI. The values of β_2 and β_0 remained similar to the superiority setting. The total sample size was estimated as $N = 676$, so that the power to detect a treatment difference when $\beta_2 = 0$ was again 80%.

It should be noted that the results on the linear/risk difference scale are not necessarily immediately applicable to the logistic/odds ratio scale but may require translation. The authors present hypotheses and parameters in terms of the linear scale to facilitate communication with clinical investigators but then analyze results using logistic regression to avoid convergence issues that would otherwise arise in the tails of the nonlinear distribution [16]. It is also noteworthy that the sample size has to be increased slightly when the effect size is translated from a risk difference to an odds ratio [20].

3.2. Simulation strategy

The simulated subject dataset is filled sequentially by first assigning the level of covariate (0 or 1), based on dichotomization of a random uniform distribution, and then a treatment indicator via either simple randomization, where the probability that the i^{th} patient is assigned to treatment ($p_{i,trt}$) is 0.5 or permuted block randomization within each level of the covariate according to the following:

$$p_{i,trt} = \frac{(b/2)\,(1 + int(i - 1/b)) - n_{i-1,trt}}{b\,(1 - int(i - 1/b)) - (i - 1)} \quad , \tag{5}$$

where i = subject, b = block size, int = next highest integer value, and n_{i-1} = number previously assigned to treatment. A block size of six was chosen as a compromise between the authors' beliefs about current popular practice and the desire for results to be comparable to those of Kahan and Morris [8], who used a block size of eight [8]. The probability of success for each patient is assigned as follows:

$$p_{i,\text{success}} = \frac{\exp(\beta_1 Trt + \beta_2 \text{Covariate})}{1 + \exp(\beta_1 Trt + \beta_2 \text{Covariate})} \quad , \tag{6}$$

and again compared with a random uniform distribution for dichotomization. Once the subject table is populated, unadjusted and adjusted analyses are conducted, and estimates of the odds ratios of the treatment effect, as well as their standard errors and two-sided 95% confidence intervals, are extracted.

Power, defined as the percentage of trials under the alternative hypothesis in which the lower bound of the confidence interval for the odds ratio crosses the predetermined margin (> 1.0 for superiority or

> 0.583 for NI), as well as type I error, defined as the percentage of trials under the null in which the lower bound of the confidence interval for the odds ratio crosses the predetermined margin, are calculated across 10,000 iterations. Bias in the estimate of the treatment effect, defined as the average difference between the estimate and the true value of β_1, and bias in the standard error of the estimate of treatment effect, defined as the mean difference between empirical and model standard errors, are calculated under the alternative hypothesis in the context of superiority and under the null hypothesis in the context of NI, as this is where one would expect to find a treatment difference.

4. Simulation results

4.1. Superiority

In the context of superiority (presented in Figure 1), there is a slight loss in power as β_2 (which represents the strength of the covariate) moves away from zero in either direction for unadjusted analyses, and a slight gain in power for adjusted analyses. Furthermore, there appears to be little impact of balancing at randomization among adjusted analyses, but some improvement of covariate adjusted permuted block randomization over simple randomization among unadjusted analyses. Type I error, under all scenarios, yields close to nominal values regardless of the value for β_2, with the exception of the scenario that employed covariate adjusted permuted block randomization coupled with an unadjusted analysis. In this scenario, type I error decreased as β_2 moved away from zero in either direction.

Under the alternative hypothesis, we can see that the unadjusted analyses' treatment effects and standard errors underestimate those of the adjusted analyses. Furthermore, the bias in the standard error appears to be less severe for covariate adjusted permuted block randomization than for simple randomization due to the correlation it creates. However, the treatment estimate, as well as its accompanying standard error, is nearly unbiased (a slight positive bias was detected but determined to be minimal) in adjusted analyses, with negligible effect of covariate balancing at randomization.

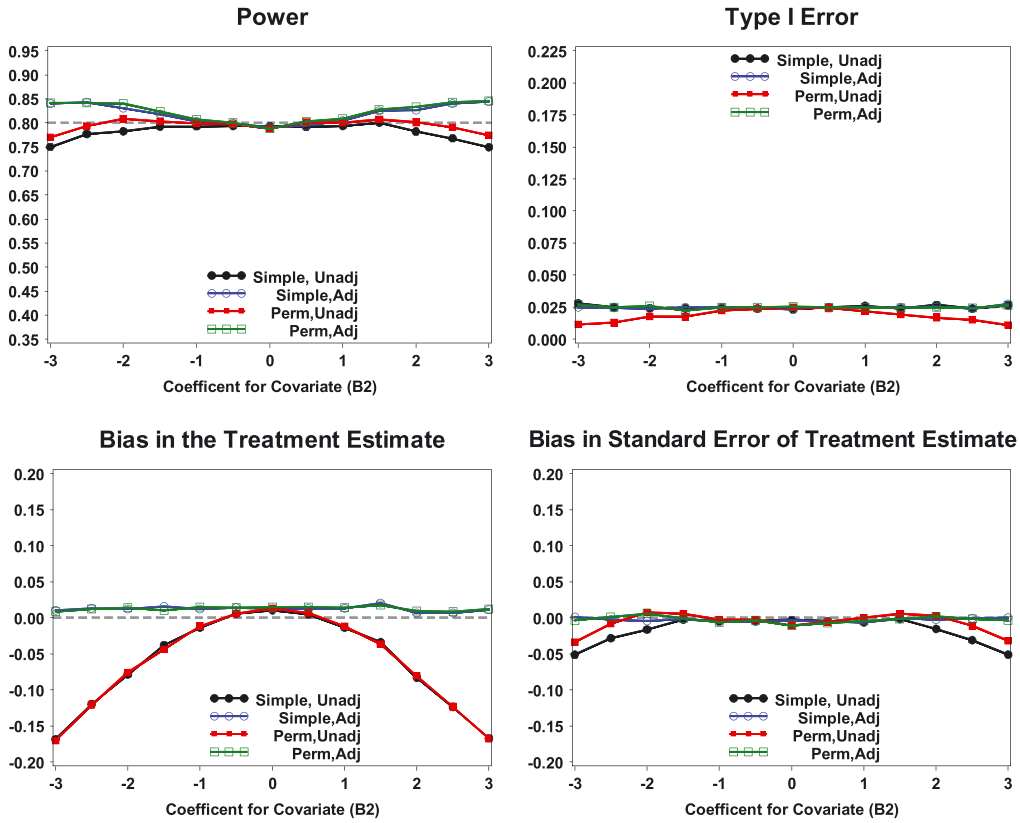

Figure 1. Superiority.

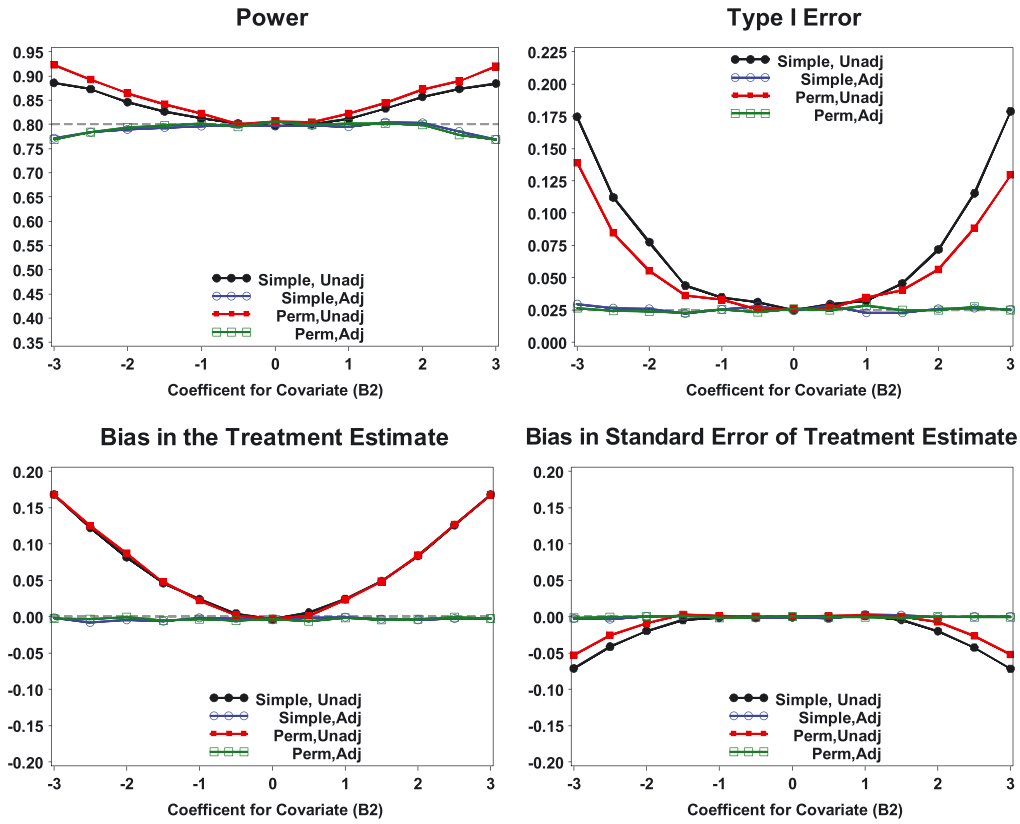

Figure 2. Noninferiority.

4.2. Noninferiority

The operating characteristics in the context of NI (presented in Figure 2) are quite different from superiority. In this setting, power decreases slightly as β_2 moves away from zero in either direction for adjusted analyses, and type I error is nearly maintained regardless of covariate balancing at randomization via permuted block. For unadjusted analyses, an opposite effect from that demonstrated in the context of unadjusted analyses in superiority is observed. Power and type I error is increased as β_2 moves away from zero in either direction. This increase in type I error for unadjusted analyses in the presence of an influential prognostic covariate is greater in simple randomization than in the context of covariate adjusted permuted block randomization. Type I error rates are nearly maintained for adjusted analyses regardless of balancing at randomization.

The treatment estimate and its standard error are unbiased for adjusted analyses, whereas the treatment estimate is biased upward toward zero for unadjusted analyses regardless of balancing at randomization via permuted block. The standard error of this estimate follows the same pattern as the alternative hypothesis in the superiority setting (namely, standard errors are deflated for unadjusted analyses with covariate adjusted permuted block randomization yielding a less pronounced effect).

5. Discussion

The results presented earlier demonstrate that adjusting for important prognostic covariates in analysis is always preferred, regardless of whether the hypothesis to be tested is one of superiority or NI. In a superiority setting, adjusted analyses yield greater power to detect a treatment difference as compared with unadjusted analyses, as well as nominal type I error. In a NI setting, adjusted analyses do not have the added benefit of increased power but are protective against unacceptable increases in type I error. Furthermore, adjusted analyses yield nearly unbiased estimates of treatment effect in both scenarios and biases in standard error that seem unavoidable given the nature of the nonlinear relationship between outcome and predictors.

These findings expand upon those of Garrett [4] by further quantifying the impact of failure to adjust for important prognostic covariates in terms of bias in the treatment estimate and subsequent implications for power and type I error. It also provides noteworthy extensions in terms of the evaluation of different covariate adjustment strategies (i.e. randomization versus analysis), a direct comparison with a superiority design and the extendibility of such findings to different specifications of parameters. For example, in addition to the strength of association between covariate and outcome, the probability of success in the control also plays a role in the variance of the treatment estimate such that different starting points do not translate to equivalent odds ratios. Thus, an absolute NI margin of 0.10, which we set in the NI setting, does not always translate to a relative NI margin of 0.583, but only given an assumed probability of success in the control of 80%. To demonstrate the sensitivity of findings, a second simulation was undertaken in which the absolute NI margin remained 0.10, but the probabilities of success were changed to $\pi_C = 0.55$ and $\pi_T = 0.45$ (which translates to a relative NI margin of 0.669). In this scenario, the overall variability is increased and the impact of failure to adjust for covariates is magnified, but similar trends were observed in all operating characteristics for the NI setting.

A limitation of this work lies in the fact that the observed differences between unadjusted and adjusted analyses can be thought of in terms of model inconsistency. The adjusted model is a conditional or subject-specific model, whereas the unadjusted model is a marginal or population averaged model. Thus, one could argue that the bias illustrated in the unadjusted treatment effect estimates is actually just a result of the difference between the two modeling strategies [21, 22]. While this is worth commenting, the simulation results, as presented, retain their practical value as an explanation of the impact of covariate adjustment.

Finally, the current research suggests that, in the context of adjusted analyses of binary outcomes, covariate adjusted permuted block randomization may not provide much gain over simple randomization. In fact, there were only minimal effects of balancing at randomization in the context of adjusted analyses for any of the operating characteristics that were presented, a result that was maintained even under a scenario in which a significant 70/30 covariate imbalance was forced (simulation not shown). In addition, the use of covariate adjusted permuted block randomization results in a loss of randomness via an increase in the number of deterministic assignments [23]. However, further research is required to determine whether these findings can be generalized to other randomization schemes.

This research illustrates the impact of covariate adjustment at randomization and analysis in NI trials with binary outcomes and demonstrates the importance of conducting adjusted analyses in the presence of important prognostic covariates. Although we have exaggerated the covariate effect (i.e. β_2), this was performed for purely theoretical purposes. In practice, it is unlikely that one would see a covariate effect of $+/-3.0$ and the resulting type I error rate of 0.175. Regardless, our results hold as we see an inflation in the type I error when the covariate is not equal to 0; for example, a more practical case of a covariate effect of $+/-1.5$ shows a type I error rate of 0.045. This suggests that a proper adjustment will result in unbiased estimates of treatment and reduce the probability of committing a type I error, which is of particular importance in NI trials, as claiming NI when it is false may have severe implications for patients.

Acknowledgements

This research is supported by a National Institute of Neurological Diseases and Stroke NINDS grant: NETT U01 NS059041. We are grateful to Patrick Mauldin and Caitlyn Ellerbe for their helpful comments and for the comments of the anonymous reviewers.

References

1. US Department of Health and Human Sciences Food and Drug Administration. Center for drug evaluation and research and center for biologics evaluation and research. FDA guidance for industry non-inferiority clinical trials, March 2010.
2. International Conference on Harmonization. Statistical principles for clinical trials (ICH E 9), Food and Drug Administration, DSSH, Febuary 1998.
3. Blackwelder WC. Proving the null hypothesis' in clinical trials. *Controlled Clinical Trials* 1982; **3**:345–353.
4. Garrett AD. Therapeutic equivalency: fallacies and falsification. *Statistics in Medicine* 2003; **22**:741–762.
5. D'Agostino RB. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. *Statistics in Medicine* 2003; **22**:169–186.
6. Durkalski V, Silbergleit R, Lowenstein D. Challenges in the design and analysis of non-inferiority trials: a case study. *Clinical Trials* 2011; **0**:1–8.
7. Dunnett CW, Gent M. Significance testing to establish equivalence between treatments, with special reference to data in the form of 2x2 tables. *Biometrics* 1977; **33**(4):593–602.

8. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine* 2012; **31**:328–340.
9. Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* 2010; **29**(2):347–360.
10. Permutt T. Adjustment for covariates. *Encyclopedia of Biopharmaceutical Statistics* 2003; **1**(1):18–21.
11. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; **8**:467–475.
12. Ciolino JD, Martin RH, Zhao W, Jauch EC, Hill MD, Palesch YY. Covariate imbalance and adjustment for logistic regression analysis of clinical trial data. *Biopharmaceutical Statistics* 2013; **23**(6):1383–1402.
13. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984; **71**(3):431–444.
14. Gail MH, Tan WY, Piantadosi S. Tests for no treatment effect in randomized clinical trials. *Biometrika* 1988; **75**(1):57–64.
15. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 1991; **59**(2):227–240.
16. Neuhaus JM, Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika* 1993; **80**(4):807–815.
17. Ford I, Norrie J, Ahmadi S. Model inconcsistency, illustrated by the cox proportional hazards model. *Statistics in Medicine* 1995; **14**:735–746.
18. Beach ML, Meier P. Choosing covariates in the analysis of clinical trials. *Controlled Clinical Trials* 1989; **10**:161S–175S.
19. Senn SJ. Testing for baseline balance in clinical trials. *Statistics in Medicine* 1994; **13**:715–1726.
20. Hilton JF. Noninferiority tial designs for odds ratios and risk differences. *Statistics in Medicine* 2010; **29**:982–993.
21. Lee Y, Nelder JA. Conditional and marginal models: another view. *Statistics Science* 2004; **19**:219–228.
22. Senn SJ. Conditional and marginal models: another view – comments and rejoinders. *Statistics Science* 2004; **19**:228–238.
23. Zhao W, Weng Y, Wu Q, Palesch Y. Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharmaceutical Statistics* 2012; **11**(1):39–48.