Published in final edited form as:

Contemp Clin Trials. 2014 May ; 38(1): 9–18. doi:10.1016/j.cct.2014.02.007.

Continuous Covariate Imbalance and Conditional Power for Clinical Trial Interim Analyses

Jody D. Ciolino^{a,*}, Renee^a H. Martin^b, Wenle Zhao^b, Edward C. Jauch^b, Michael D. Hill^c, and Yuko Y. Palesch^b

^aDepartment of Preventive Medicine, Northwestern University, Chicago, IL

^bMedical University of South Carolina, Charleston, SC, USA

^cUniversity of Calgary, AB, Canada

Abstract

Oftentimes valid statistical analyses for clinical trials involve adjustment for known influential covariates, regardless of imbalance observed in these covariates at baseline across treatment groups. Thus, it must be the case that valid interim analyses also properly adjust for these covariates. There are situations, however, in which covariate adjustment is not possible, not planned, or simply carries less merit as it makes inferences less generalizable and less intuitive. In this case, covariate imbalance between treatment groups can have a substantial effect on both interim and final primary outcome analyses. This paper illustrates the effect of influential continuous baseline covariate imbalance on unadjusted conditional power (CP), and thus, on trial decisions based on futility stopping bounds. The robustness of the relationship is illustrated for normal, skewed, and bimodal continuous baseline covariates that are related to a normally distributed primary outcome. Results suggest that unadjusted CP calculations in the presence of influential covariate imbalance require careful interpretation and evaluation.

Keywords

covariate adjusted analysis; conditional power; covariate imbalance

1 Introduction

In a randomized setting, clinical trial treatment arms will be comparable on average with respect to covariate distributions. Thus, the expected level of covariate imbalance in a randomized clinical trial is zero and adjusted and unadjusted analyses will generally result in the same overall conclusions, but observed imbalance in a single clinical trial is one realization of all possible random levels of imbalance. Therefore, a single clinical trial may

© 2014 Elsevier Inc. All rights reserved.

*Corresponding author: Jody D. Ciolino, Department of Preventive Medicine, Northwestern University, 680 N Lake Shore Drive Suite 1400, Chicago, IL 60611, jody.ciolino@northwestern.edu (Jody D. Ciolino).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

exhibit some form of nontrivial covariate imbalance for which adjustment should be made in analyses. The statistical literature argues that adjustment is essential in clinical trial analysis for known influential baseline covariates in order to ensure statistical efficiency and unbiased treatment effect estimates¹⁻⁷, but according to International Conference on Harmonization (ICH) guidelines, adjustment in statistical analysis for covariates known to affect primary outcome must be pre-specified in the trial's statistical analysis plan (SAP)⁸. Unplanned adjusted analyses are thus considered secondary and carry less merit than planned unadjusted primary analyses.

However, choosing covariates to include in a final statistical model for primary outcome can be a difficult task for clinicians and statisticians designing clinical trials as situations arise in which influential covariates are unknown ahead of time. For example, the original analysis of the National Institute of Neurological Disorders and Stroke (NINDS) tissue plasminogen activator (tPA) study for ischemic stroke⁹ failed to account for baseline NIH Stroke Scale (NIHSS) score, a measure of baseline disease severity, resulting in controversy surrounding the efficacy of tPA in the treatment of ischemic stroke as well as the need for reanalysis of these data¹⁰⁻¹³.

Hauck², Hernandez³, and Peduzzi et al.⁴ discuss the ease of interpretation and generalizability of unadjusted treatment effect estimates when compared to the adjusted estimates based on models. As a result, clinical trial reports often place emphasis on simple, unadjusted treatment effect estimates^{2;3;14}. Austin et al.¹⁴ suggest that a larger percentage of unadjusted analysis results are reported in clinical trial articles when compared to adjusted results. Although validity of analysis in the presence of known influential covariates requires proper adjustment¹⁵, balance in influential baseline covariates may serve as a compromise between the complex, but more appropriate adjusted analyses and the more readily interpretable and accepted unadjusted analysis in some cases.

Senn¹⁶ shows that imbalance in continuous normal covariate distributions across treatment groups as measured by the Z- or t-statistic is directly associated with type I error inflation in an unadjusted analysis for continuous primary outcome. Ciolino et al.^{17;18} illustrate the robustness of the t-statistic in predicting power, type I error rate, and bias in unadjusted analyses for several continuous covariate distributions. The impact of imbalance on these statistical parameters for unadjusted analyses depends on the level of covariate influence on primary outcome.

If imbalance in continuous baseline covariate distributions is predictive of statistical parameters in the final analysis, then such is the case for an interim point of a trial and this imbalance may thus indirectly affect trial decisions based on unadjusted interim analyses. For example, the data monitoring committee (DMC) for a clinical trial may decide to terminate the trial prematurely because conditional power (CP) at an interim point falls below a pre-specified stopping boundary¹⁹. Presence of covariate imbalance in the case of unadjusted CP can therefore potentially have an indirect effect on the DMC's decision to terminate enrollment. This paper aims to determine the relationship between continuous baseline covariate imbalance and unadjusted CP at interim analysis for a normally distributed primary outcome and to illustrate the potential effect this imbalance may have on

trial decisions based on unadjusted CP. We argue that unadjusted CP calculations require careful interpretation, and one should consider calculating CP based on a test statistic adjusted for influential and/or imbalanced covariates.

2 Background

2.1 Statistical Framework

The basic statistical ideas outlined here adopt those of Lan and Wittes²⁰ and Lan et al.²¹ for CP and Brownian Motion properties. Consider a randomized clinical trial with two arms: an active treatment group and a placebo group with 1:1 allocation. Assume n out of the total N subjects have been enrolled in each arm, and let $t = n/N$ denote the trial fraction at an interim point. Further, assume the primary outcome, $Y \sim N(\mu_Y, \sigma_Y^2)$, and planned analysis fails to adjust for an influential covariate, $X \sim N(\mu_X, \sigma_X^2)$.

Let $Z_X(t) = \frac{\mu_{X_{tx}} - \mu_{X_{pbo}}}{\sigma_X \sqrt{2/n}}$ (where $\mu_{X_{tx}}$ is the mean covariate value for the active treatment group and $\mu_{X_{pbo}}$ is the mean covariate value for the placebo group) and $Z_Y(t) = \frac{\mu_{Y_{tx}} - \mu_{Y_{pbo}}}{\sigma_Y \sqrt{2/n}}$ (where $\mu_{Y_{tx}}$ is the mean primary outcome value for the active treatment group and $\mu_{Y_{pbo}}$ is the mean primary outcome value for the placebo group) represent the Z-scores at trial fraction t comparing mean covariate and primary outcome values across treatment groups, respectively. Following the notation of Lan and Wittes²⁰, the B-values used in calculating CP are equivalent to $B_X(t) = \sqrt{t}Z_X(t)$ and $B_Y(t) = \sqrt{t}Z_Y(t)$, respectively.

Let θ be the expected Z-score for the primary outcome at the end of the trial (i.e., $\theta = E[Z_Y(1)]$). Under random allocation, treatment groups are expected to be balanced with respect to continuous covariates, and the expected Z-score for the covariate at the end of the trial is zero (i.e., $E[Z_X(1)] = 0$). Assume that $\theta > 0$ such that the treatment has a positive effect on primary outcome.

It should be noted that in the calculation of $Z_X(t)$ and $Z_Y(t)$, the numerators are calculated in the same direction (mean value in the treatment group minus mean value in the placebo group), and the results to follow rely heavily on this fact. In the hypothetical clinical trial scenario discussed here, assume that larger values of outcome correspond to more favorable clinical prognosis and situations in which the placebo group is “favored” at baseline suggest a better baseline prognosis in the placebo group (i.e., the placebo group is predisposed for better clinical outcome).

It can be shown that if the $\text{corr}(X, Y) = \rho$, then the $\text{corr}(B_X(t), B_Y(t)) = \rho$. Given this information and the properties of Brownian Motions (as related to the B-values here), we can determine the distribution of $B_Y(1)|B_X(t) = b_{X_t}$. This distribution can, in turn, be used to calculate unadjusted CP given only covariate imbalance for an influential covariate at trial fraction t .

2.2 The Distribution of $B_Y(t)|B_X(t) = b_{X_t}$

The properties of Brownian Motions and their relationship to the B-value in clinical trial data monitoring²⁰ allow for the following assumptions:

- $B_X(t) \sim N(0, t)$
- $B_Y(t) \sim N(\theta t, t)$
- $\text{corr}(B_X(t), B_X(1)) = \text{corr}(B_Y(t), B_Y(1)) = \sqrt{t}$
- $\text{corr}(B_X(t), B_Y(t)) = \rho$

By the definition of the conditional normal distribution, it can be shown that

$$\begin{aligned} (B_X(1)|B_X(t)=b_{X_t}) &\sim N(b_{X_t}, 1-t), \\ \text{and} \\ (B_Y(1)|B_X(1)=b_{X_1}) &\sim N(\theta + \rho b_{X_1}, 1-\rho^2). \end{aligned}$$

However, $B_X(1)$ is a random variable that depends on $B_X(t) = b_{X_t}$. Therefore,

$$E(B_Y(1)|B_X(t)=b_{X_t}) = \theta + \rho b_{X_t}.$$

Using this fact and the properties of conditional normal distributions, it can be seen that

$$\begin{aligned} \text{corr}(B_Y(1), B_X(t)) &= \rho \sqrt{t}, \\ \text{and thus,} \\ \text{Var}(B_Y(1)|B_X(t)=b_{X_t}) &= 1 - \text{corr}(B_Y(1), B_X(t))^2 = 1 - t\rho^2. \end{aligned}$$

Therefore,

$$(B_Y(1)|B_X(t)=b_{X_t}) \sim N(\theta + \rho b_{X_t}, 1 - t\rho^2). \quad (1)$$

2.3 Relationship between unadjusted CP and Covariate Imbalance

Equation (1) may be used to calculate the CP (i.e., the probability of rejecting the null hypothesis of no treatment effect at the end of the trial) given a value of θ and current baseline covariate imbalance (defined by the Z-statistic comparing mean covariate values across treatment groups). Assume the null hypothesis (H_0) for an unadjusted test on primary outcome is that of no treatment effect against a one-sided alternative, such that one would reject H_0 if $Z_Y(1) = B_Y(1) \geq Z_\alpha$. Using this rejection rule, CP calculated with the unadjusted test statistic conditional only on covariate imbalance is equivalent to

$$CP_{Z_x} = 1 - \Phi \left[\frac{Z_\alpha - (\theta + \rho \sqrt{t} Z_x(t))}{\sqrt{1 - t\rho^2}} \right], \quad (2)$$

where Φ represents the cumulative distribution function for the standard normal. Again, note that the subtraction expression in the numerator calculations for Z-statistics must be in the same direction (i.e., mean value in treatment group minus mean value in the placebo group) for both primary outcome and covariate.

Note that the purpose of this derivation is not to provide a replacement formulation of CP, but it is meant to illustrate the fact that CP for an unadjusted analysis depends on the level covariate imbalance at trial fraction t and the correlation of that covariate with outcome. It can be shown that $B_Y(1)|(B_X(t), B_Y(t)) = (b_{X_t}, b_{Y_t})$ is equivalent to

$$(B_Y(1)|B_Y(t)=b_{Y_t}) \sim N(b_{Y_t} + \theta(1-t), 1-t). \quad (3)$$

This relationship is used to calculate CP in the traditional sense, and as previously stated, Equations (1) and (2) are not presented here to replace this formulation. It is true that unadjusted $B_Y(t)$ depends on $B_X(t)$. Equations (1) and (2) illustrate the relationship of unadjusted CP conditional only on current imbalance ($B_X(t)$).

Just as in the traditional sense, CP_{Z_X} may be calculated under the null hypothesis ($\theta = 0$), under the alternative hypothesis ($\theta = \delta / (\sigma_y \sqrt{2/N})$, where δ is the hypothesized treatment effect), or under the current trend of the data ($\hat{\theta} = b_{Y_t} / t = Z_Y(t) / \sqrt{t}$). Figure 1 illustrates the relationship between covariate imbalance (measured by the Z-statistic) and CP_{Z_X} at trial fraction $t = 0.60$ under the alternative hypothesis that corresponds to 80% power for an unadjusted analysis (using Equation (2)). A positive imbalance in this case suggests that the placebo group is favored at baseline with respect to the covariate (correlation between covariate and outcome is actually negative, but for ease of interpretation, the sign of correlation coefficients has been omitted). Intuitively, if the placebo group is favored at baseline, then unadjusted analysis would not detect the true treatment effect as easily as it would under adequate baseline covariate balance. The larger the level of association between outcome and covariate, the greater the effect of imbalance on unadjusted CP (Figure 1).

Simulated data are required in order to illustrate the relationship between CP under the current trend and baseline covariate imbalance at an interim point. Five thousand (5000) clinical trials were simulated involving two arms, a normally distributed outcome, and a normally distributed covariate with varying levels of covariate influence ($\rho = 0, 0.3, 0.6, 0.8$). Treatment effect was simulated to correspond to 80% power for an unadjusted analysis on the primary outcome. CP (for unadjusted analysis, using the relationship depicted in Equation (3)) under the current trend was calculated in each simulated trial after 60% of subjects had been enrolled. Figure 2 illustrates the relationship between unadjusted CP under the current trend and imbalance (measured by the t-statistic comparing mean covariate

values across treatment groups) using the locally weighted least squares smoothing for the simulated data.

Figure 2 shows a similar relationship to the one seen for CP_{Z_x} calculated under the alternative hypothesis, but the effect of imbalance on unadjusted CP under the current trend (Figure 2) is more substantial. Note that a correlation between covariate and outcome as high as 0.80 is seldom observed in practice (with the exception of the baseline value of outcome). The simulated levels of association were selected for illustrative purpose so as to show the nature of the relationship in extreme ($\rho = 0, 0.8$) and also the moderate ($\rho = 0.30, 0.60$) cases. Impact of specific levels of association (e.g., $\rho = 0.215$) may be seen through appropriate substitution in Equation (2).

Determining whether to adjust analyses (interim or final) for an influential covariate can be a difficult process for both clinicians and statisticians. Often-times, investigators will assume that if there are no “statistically significant” differences between treatment groups with respect to influential covariates, there is no need to adjust for these covariates in analyses. Several authors^{14–18} have argued against this practice. Figures 1 and 2 provide further evidence, indicating that statistically insignificant levels of imbalance do not justify failing to adjust for influential covariates when calculating CP.

These figures each include a vertical line at the value of 1.96 for reference. This line corresponds to the two-sided 5% level of significance in a baseline test comparing mean covariate values across treatment groups. In each figure (especially Figure 2), the effects of covariate imbalance on unadjusted CP are nontrivial well before a “statistically significant” level of baseline covariate imbalance is observed. Recall that simulation assumed a treatment effect corresponding to 80% power in each simulated trial. Thus, a low unadjusted CP may lead to incorrect decisions to terminate the trial, and imbalance in influential covariates may be the reason for such mistakes. For example, consider a pre-specified futility stopping rule that states a given clinical trial will be stopped for futility if a CP calculation based on the current trend at trial fraction $t=0.60$ falls below 20%. If this interim analysis does not account for a highly influential baseline covariate ($\rho = 0.80$), then according to Figure 2, an imbalance in that covariate corresponding to a two-sided 13% level of significance ($t=1.5$) may result in an incorrect decision to terminate the clinical trial after only 60% of subjects have enrolled.

Derivation of Equation (2) assumes a normally distributed outcome and a normally distributed covariate, and it assumes that analysis is unadjusted for the influential covariate. The true level of covariate influence as measured by ρ may not be known, and the sample Pearson correlation coefficient would serve as an appropriate estimate. Additionally, the true standard deviations for the outcome (Y) and covariate (X) will most likely be unknown. Therefore, the Z-statistics, $Z_Y(t)$ and $Z_X(t)$, may also require estimation with the respective two-sample independent t-statistics (Figure 2). The next sections will outline computer simulation methods, and results suggest the t-statistic is relatively robust (omitted here) when inserted into Equation (2) in place of the Z-statistic. The simulations also examined the scenarios in which the assumption of normality for the covariate (X) is violated, and they

were further used to determine whether a relationship exists between covariate imbalance and CP when interim analyses are properly adjusted for the covariate of interest.

Figures 1 and 2 illustrate the effect of interim covariate imbalance that “favors” the placebo group, resulting in a decrease in calculated CP. It may be the case that the imbalance occurs in the opposite direction (randomly or possibly due to selection bias) resulting in a treatment group with a more favorable disposition at baseline. This may result in an unwarranted increase in unadjusted CP at a given time point which would affect the decision to continue or stop a clinical trial based on pre-specified futility stopping bounds. The magnitude of this effect can be quantified using Equation (2). Refer to Sections 3.2 and 4.3 for further exploration of this topic.

3 Methods

Five thousand (5000) clinical trials involving two treatment arms, a normally distributed outcome and a continuous baseline covariate were simulated under several scenarios. The scenarios included situations of a highly skewed baseline covariate (lognormally distributed), a normally distributed covariate, and a bimodally distributed covariate. Sample sizes of 100, 300, 500, and 1000 were explored along with varying levels of association ($\rho = 0, 0.3, 0.6, 0.8$) between covariate and outcome. Simulations assumed a treatment effect corresponding to that which would be observed under 80% power. In each simulated clinical trial, adjusted and unadjusted CP was calculated at pre-specified interim points. These simulations used the CP calculation based on the current trend of the data as outlined by Lan and Wittes²⁰ and Lan et al.²¹ at two interim points (trial fractions 0.30 and 0.60). Adjusted CP was calculated based on test statistics for analyses that properly adjusted (using analysis of covariance or ANCOVA) for the influential covariate of interest. Note that in practice, CP should generally be calculated based on the null, alternative, and current trend. All three results should carry weight in decisions regarding futility.

3.1 The Effect of Imbalance on Stopping Rules

If an interim CP calculation results in an estimate that falls below a pre-specified threshold, the trial's DMC may suggest termination of trial recruitment for futility¹⁹. Recall that simulations assumed a treatment effect corresponding to 80% power in each scenario, so ideally, CP calculations would not result in decisions to terminate recruitment in these simulations. The simulated data in each scenario were used to determine the probability of incorrectly stopping a clinical trial based on the unadjusted CP estimate falling below a specific value (10%, 15%, 20%) given influential covariate imbalance. For example, if the futility stopping rule was to stop an individual clinical trial if CP at an interim point fell below 20%, then an indicator variable was created for each of the 5000 simulated clinical trial CP calculations. If an individual CP fell below 20%, then the new indicator variable was given the value 1. Otherwise, the variable was given the value 0. Then all simulated data were used to construct a generalized linear model for this indicator variable using logit link with the covariate imbalance (t-statistic comparing mean covariate values across treatment groups) predictor.

3.2 The Effect of Imbalance on Continuation of Clinical Trials

Just as the DMC of a clinical trial may incorrectly decide to stop a clinical trial for futility, the committee also may incorrectly decide to continue the clinical trial if CP fails to fall below the pre-specified futility threshold. To explore this situation, the same simulations were conducted simulating no treatment effect for each of the scenarios previously outlined. Thus at an interim point, one would hope that the CP would fall below the pre-specified futility threshold, resulting in the decision to stop the clinical trial for futility. The simulated data were used to estimate the probability of incorrectly continuing a clinical trial based on the unadjusted CP estimate failing to fall below a specific value (10%, 15%, 20%) given covariate imbalance. For example, if the futility stopping rule was to stop an individual clinical trial if CP at an interim point fell below 20%, then a new indicator variable was created for each of the 5000 simulated clinical trial CP calculations. If an individual CP remained above 20%, then the new indicator variable was given the value 1. Otherwise, the variable was given the value 0. Then a generalized linear model for this new indicator variable was constructed using logit link with the covariate imbalance (t-statistic comparing mean covariate values across treatment groups) as a predictor.

3.3 Resampling with Replacement from the NINDS tPA Dataset

Bootstrap simulations involving sampling with replacement from the NINDS tPA dataset were also conducted. The simulations explored two primary outcomes; the first was NIHSS score at three months and the second was “successful” outcome at three months as determined by modified Rankin Score (mRS) of 0 or 1. The simulations employed the following logic:

1. If N represents the total number of subjects in the study (N=624), sample N/2 subjects with replacement such that an equal number of subjects are allocated to each treatment group. This simulates a trial fraction of 0.50. The relevant information from each of the sampled subjects includes the two outcomes, baseline NIHSS, and treatment assignment (active or placebo).
2. Conduct an unadjusted hypothesis test for treatment effect and calculate CP based on the test statistic for each outcome (t-test is used for NIHSS and the chi-squared test for binomial proportions is used for binary mRS outcome) and the current trend. Note that some clinicians may not view the NIHSS score as “continuous;” however, for the purposes of these simulations it was treated as such.
3. Conduct an adjusted (for baseline NIHSS score) analysis for each outcome and calculate CP based on the test statistic associated with treatment effect (ANCOVA is used for the NIHSS outcome, and logistic regression is used for binary mRS outcome) and the current trend.
4. Calculate covariate (baseline NIHSS) imbalance as measured by the t-statistic comparing mean covariate values across treatment groups.
5. Return to step 1.

Note that a positive treatment effect was observed (whether adjusted or unadjusted for baseline NIHSS) in the tPA dataset, and baseline NIHSS has a significant negative

association with each outcome (larger baseline NIHSS values are generally associated with poorer prognosis at three months [i.e., higher three-month NIHSS scores, and higher probability of negative outcome]). Therefore, it was expected that an imbalance in baseline NIHSS such that the treatment group has poorer baseline prognosis (positive imbalance) would have a negative impact on CP.

4 Results

4.1 Unadjusted vs. Adjusted CP

In the simulations explained above, imbalance in influential covariates as measured by the t-statistic was highly predictive of CP when it was calculated based on unadjusted analysis ($p < 2 \times 10^{-16}$). However, when analysis properly adjusted for the influential covariate of interest, this imbalance was never (in any of the scenarios examined) predictive of CP calculated based on the appropriate test statistic.

Figure 3 shows the bias (adjusted analysis CP minus unadjusted analysis CP) of the CP calculation for unadjusted analysis given covariate influence and imbalance. When the covariate had no influence on primary outcome, there was no observed association between covariate imbalance and bias in calculating unadjusted CP. However, as the level of covariate influence increased, the relationship between covariate imbalance and bias in unadjusted CP estimation became obvious. When the active treatment group was favored (imbalance was negative, t-statistic < 0), the unadjusted CP overestimated the appropriately adjusted CP estimates. On the other hand, when the placebo group was favored (imbalance was positive, t-statistic > 0), the unadjusted CP underestimated the appropriately adjusted CP estimates. Furthermore, as the magnitude of imbalance increased in either direction, the magnitude of bias in unadjusted CP increased. When imbalance was essentially zero, the bias in unadjusted CP estimates was also essentially zero. These results were consistent for different covariate distributions (normal, lognormal, and bimodal) and across all sample sizes explored.

4.2 The Effect of Imbalance on Stopping Rules

In all scenarios, when the correlation between covariate and outcome was nonzero, the current covariate imbalance was a significant predictor of whether the simulated trial would be stopped ($p < 2 \times 10^{-16}$ in almost all cases). Figure 4 shows the estimated probability of incorrectly stopping (based on the generalized linear models from the simulated data) a clinical trial at trial fraction 0.30 given current covariate imbalance. The scenarios shown in Figure 4 correspond to sample sizes of 1000 and a simulated skewed or lognormal covariate distribution. The results were consistent across sample size as well as covariate distribution. The imbalance seemed to have no effect on whether a trial would be stopped for futility when there was no simulated association between covariate and outcome. However, as the level of covariate influence increased, the amount of influence covariate imbalance had on the decision to stop the trial for futility also increased (Figure 4). Each plot contains a vertical line at 1.96 to serve as reference for a “statistically significant” level of imbalance (at the 5% level).

According to these results, the probability of stopping a clinical trial for futility based on unadjusted CP and these stopping rules is nontrivial well before covariate imbalance reaches the “statistically significant” level. In fact, when covariate imbalance for a moderately influential covariate ($\rho = 0.60$) corresponds to an 11% (two-sided) significance level (imbalance = 1.58), there is an estimated 50% chance of a stopping a clinical trial for futility based on unadjusted CP at just 30% of the way through the trial. In addition, when imbalance corresponds to almost a 20% significance level (imbalance = 1.27) for a highly influential ($\rho = 0.80$) covariate, there is also an estimated 50% chance of stopping a clinical trial for futility based on unadjusted CP after 30% of the total subjects have been enrolled.

4.3 The Effect of Imbalance on Continuation of Clinical Trials

In all scenarios, when the correlation between covariate and outcome was nonzero, the current covariate imbalance was a significant predictor of whether the trial would continue ($p < 2 \times 10^{-16}$ in almost all cases). Figure 5 shows the estimated probability of incorrectly continuing (based on the generalized linear models from the simulated data) a clinical trial at trial fraction 0.60 given current covariate imbalance for various stopping rules. In Figure 5, a positive imbalance (t-statistic > 0) means that the treatment group was favored at baseline with respect to the covariate. This is different from previous figures, where a positive imbalance meant that the placebo group was favored at baseline. The scenarios shown in Figure 5 correspond to sample sizes of 1000 and a simulated skewed or lognormal covariate distribution. The results were relatively consistent across sample size as well as covariate distribution. As is evidenced by Figure 5, the imbalance seemed to have no effect on whether a trial would continue enrolling subjects when there was no simulated association between covariate and outcome. However, as the level of covariate influence increased, the amount of influence covariate imbalance had on the decision to continue the trial for lack of futility (based on unadjusted CP) also increased. Each plot contains a vertical line at 1.96 to serve as reference for a “statistically significant” level of imbalance (at the 5% level).

Table 1 uses the simulation results to summarize the probabilities of making incorrect decisions at an interim point (both trial fractions 0.3 and 0.6 are shown) based on unadjusted CP calculations for given levels of covariate imbalance. Note that the levels of imbalance in Table 1 (measured by the t-statistic or Z-statistic) correspond to approximate 20%, 10%, and 5% (two-sided) levels of significance. When the correlation between covariate and outcome is zero, the probability of making incorrect decisions seems unaffected by covariate imbalance. However, as the level of covariate influence increases, the impact of imbalance on the probability of making incorrect decisions also increases. Note that the direction of imbalance is important in determining its effect. Underestimation of CP (corresponding to the P(Stop) columns) occurs when the placebo group is favored at baseline, and overestimation of CP (corresponding to the P(Continue) columns) occurs when the treatment group is favored at baseline with respect to the influential covariate. For ease of interpretation, Table 1 simply reports the magnitude of imbalance, but the directionality is essential in determining this effect on CP. The next section summarizes results from resampling from the NINDS tPA dataset.

4.4 Resampling with Replacement from the NINDS tPA Dataset

In the bootstrap samples, baseline NIHSS imbalance was significantly negatively correlated with unadjusted CP calculations at trial fraction 0.50 ($p < 2 \times 10^{-16}$). That is, imbalance favoring the placebo group at baseline (imbalance > 0) was negatively associated with CP. Pearson sample correlation coefficients for CP versus baseline NIHSS score in the bootstrap simulations were -0.45 (-0.48, -0.43) and -0.34 (-0.36, -0.31) for the NIHSS and mRS outcomes at three months, respectively.

Baseline NIHSS imbalance was not significantly associated with CP estimates based on adjusted test statistics in the bootstrap simulations. When plotting unadjusted CP bias (adjusted CP minus unadjusted CP) against baseline NIHSS score imbalance in the bootstrapped data, similar plots to those seen in Figure 3 (for nonzero correlations) were observed (plot omitted here). As the magnitude of baseline NIHSS imbalance strayed from zero, the magnitude of CP underestimation (or overestimation) for the unadjusted cases increased.

In addition, when using the bootstrap samples to model the probability of stopping the trial given a level of baseline NIHSS imbalance, the results were similar to those seen for the simulated scenarios, and the shape of the plots of probability of stopping the trial based on a pre-specified futility rule against NIHSS imbalance at baseline closely resembled those seen in Figures 4 and 5. Note that it is not possible to determine whether the decision to stop the trial would have been “incorrect” as with the simulated hypothetical scenarios because the underlying relationship between the treatment and outcome is not truly known in the NINDS tPA dataset. Nonetheless, the relationship between the probability of stopping the trial for futility and baseline covariate imbalance (NIHSS score) remains evident in the bootstrap simulations from the NINDS dataset.

5 Discussion

In clinical trial data analysis, it is imperative to adjust treatment effect estimates for covariates known to influence primary outcome^{1-7,16}. Failure to do so may result in inflated type I error rates^{15,16}, biased treatment effect estimation²², or loss in statistical efficiency (i.e., power to detect a true treatment effect)^{1-3,5-7}. If this is true for final primary outcome analysis, this must also apply to analyses at the interim stages of clinical trials. Equation (2) illustrates this point as it shows the relationship between influential covariate imbalance at a given trial fraction and unadjusted CP under several assumptions. Ciolino et al.^{17,18} have illustrated the robustness of the t-statistic in measuring covariate imbalance when compared to several other statistics, including nonparametric measures such as the Wilcoxon Rank-Sum statistic, Kolmogorov-Smirnov statistic, and the area under the curve of covariate imbalance across two treatment groups.

Since the relationship between unadjusted CP and covariate imbalance as measured by the t-statistic has potential to be nontrivial, it is necessary to adjust for influential covariate(s) in calculation of test statistics for primary outcome used in CP calculations. Although CP should not be the only factor driving a decision to discontinue a trial, it is one tool that DMCs often utilize¹⁹ to assess evidence of futility and aid in major trial decisions. Thus,

properly adjusted CP calculation is essential, and failure to account for continuous covariates in interim analysis may result in an incorrect decision to terminate early (or to continue) that is directly related to the level of imbalance observed in that covariate (Figures 4 and 5). Therefore, unadjusted CP must be interpreted with caution in the presence of influential covariates.

Furthermore, “insignificant” levels of imbalance do not suggest continuous covariate balance across treatment groups, and they do not justify failure to adjust for these covariates if they are known to influence primary outcome. For example, Figure 3 shows that unadjusted CP can underestimate the adjusted CP by 50% in some cases well before covariate imbalance reaches a nominal 5% level of significance. In addition, Figures 4 and 5 show that the probability of making incorrect trial decisions based on unadjusted conditional power becomes nontrivial before covariate imbalance reaches the nominal 5% level.

In a debate on futility analysis²³, Meade argues against the use of CP calculations in decisions to terminate a clinical trial for futility. Her argument focuses on the ALVEOLI trial that examined higher versus lower positive end-expiratory pressure (PEEP) in treating patients with acute respiratory distress syndrome (ARDS)²⁴. This trial was stopped after 549 of the projected 750 subjects were enrolled because the calculated conditional power fell below a pre-specified futility stopping boundary. In fact, probability of observing a significant result at the end of the trial was less than 1% based on an *unadjusted* mortality difference. Meade mentions that two important baseline covariates were imbalanced across treatment groups: age and disease severity, and proper adjustment for these variables actually shows a positive result at the interim^{23;24}. Though this reversal of treatment effect direction is not statistically significant, Meade argues that continuing the trial would have made things more clear²³. The problem with the CP calculation for the ALVEOLI trial was it failed to take into account influential covariates that were imbalanced across treatment groups. As a result, an unadjusted CP may have been the reason for a possibly incorrect decision to terminate trial enrollment. Meade uses this as an argument against the use of CP in making trial decisions at the interim. However, the results illustrated in this paper suggest that argument should not be made against CP calculation altogether, but that interim analyses should strive for *valid* CP calculations that properly adjust for known influential covariates.

It is suggested under ICH guidelines⁸ that any adjusted analyses be planned ahead of time in the SAP. However, situations may arise in which influential covariates may not be known ahead of time but discovered throughout the course of the trial, and as a result, adjustment for these variables may not have been planned a priori. Historically there has been more emphasis on unadjusted treatment effect estimates due to their ease of interpretation and generalizability²⁻⁴. For all of these reasons, unadjusted statistical analyses (whether at the interim or at the final analysis stage) may carry more weight in inference and interpretation. Therefore, a compromise between these less statistically valid unadjusted analyses and the more complicated, but appropriate adjusted analyses may be covariate balance. If influential covariates are known ahead of time, balance across treatment groups may be accomplished at the design phase through covariate adaptive treatment allocation schemes. If covariates are unknown prior to trial commencement, data integrity may be evaluated at the interim (as

seen here) or at the end of the trial (as illustrated by the work of Ciolino et al.^{17,18}) based on imbalance observed for discovered covariates. Nonetheless, a baseline test for significance (especially at the 5% level of significance) cannot be used to evaluate imbalance in covariate distributions across treatment groups.

In the case of interim analyses, unadjusted CP should not be the sole determining factor in a decision to terminate a clinical trial prematurely. As the results presented in this manuscript suggest, if influential covariates are known or become known at the interim stages of a clinical trial, CP calculation should take the adjusted treatment effect estimates into account, even if adjustment was not planned ahead of time. If this is not possible, then covariate balance decreases bias in unadjusted CP calculation, thereby decreasing the probability of making incorrect crucial trial decisions. Therefore, these results (Equation (2), Figures 3-5, and Table 1) may serve as a guide for evaluation of the reliability of unadjusted CP estimates in the presence of continuous baseline covariate imbalance.

This work assumes a normally distributed continuous outcome; however, additional research should explore the relationship between unadjusted CP and covariate imbalance for additional outcome types. Ciolino et al.¹⁸ have explored the relationship between statistical parameters in final binary primary outcome analyses and continuous covariate imbalance. One can use these results along with the results from the current paper in order to make inferences regarding CP and covariate imbalance in situations involving a non normal outcome. It should be noted that a simple correlation coefficient (ρ) cannot be used to explore the relationship between a binary outcome and continuous covariate as in the current paper. Ciolino et al.¹⁸ utilize the “ β ” term from the logisitic regression model relating outcome to covariate that corresponds to the effect of a single unit increase in the covariate on log(odds) of primary outcome to measure this association. In addition, future work should explore the impact of covariate imbalance for more than one covariate and higher order covariate terms (i.e., interactions and/or quadratic terms), and future work will also involve exploration of the performance of the most commonly used allocation techniques (i.e., stratification and dynamic allocation) in achieving relative covariate balance in order to prevent such detrimental effects on unadjusted statistical analyses at the interim and final stages of a clinical trial.

References

1. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Statist Med.* 2002; 21:2899–2908.
2. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials.* 1998; 19:249–256. [PubMed: 9620808]
3. Hernandez AV, Streyerberg EW, Habbema DF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology.* 2004; 57:454–460. [PubMed: 15196615]
4. Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. *Epidemiologic Reviews.* 2002; 24:26–38. [PubMed: 12119853]
5. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika.* 1984; 71:431–444.
6. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review.* 1991; 58:227–240.

7. Raab GM, Day S. How to select covariates to include in the analysis of a clinical trial. *Controlled Clinical Trials*. 2000; 21:330–342. [PubMed: 10913808]
8. International Conference on Harmonization E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonized tripartite guideline. *Statist Med*. 1999; 18:1905–1942.
9. NINDS. The national institute of neurological disorders and stroke rt-PA stroke study group: Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine*. 1995; 333:1581–1587. [PubMed: 7477192]
10. Ciolino JD, Zhao W, Martin RH, Palesch YY. Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization. *Contemporary Clinical Trials*. 2010; 32:250–259. [PubMed: 21078415]
11. Frey JL. Recombinant tissue plasminogen activator (rtPA) for stroke: The perspective at 8 years. *The Neurologist*. 2005; 11:123–133. [PubMed: 15733334]
12. Ingall TJ, O'Fallon WM, Asplund K, Goldfrank LR, Hertzberg VS, Louis TA, et al. Findings from the reanalysis of the ninds tissue plasminogen activator for acute ischemic stroke trial. *Stroke*. 2004; 35:2418–2424. [PubMed: 15345796]
13. Hertzberg VS, Ingall TJ, O'Fallon WM, Asplund K, Goldfrank LR, Louis TA, et al. Methods and processes for the reanalysis of the ninds tissue plasminogen activator for acute ischemic stroke treatment trial. *Clinical Trials*. 2008; 5:308–315. [PubMed: 18697845]
14. Austin PC, Manca A, Zwarenstein M, Juurlink DN, Stanbrook MB. A substantial and confusing variation exists in the handling of baseline covariates in randomized controlled trials: A review of trials published in leading medical journals. *Journal of Clinical Epidemiology*. 2010; 63:142–153. [PubMed: 19716262]
15. Senn S. Testing for baseline balance in clinical trials. *Statist Med*. 1994; 13:1715–1726.
16. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statist Med*. 1989; 8:467–475.
17. Ciolino JD, Martin RH, Zhao W, Hill MD, Jauch EC, Palesch YY. Measuring continuous baseline covariate imbalances in clinical trial data. *Statistical Methods in Medical Research*. 2011; 10.1177/0962280211416038
18. Ciolino JD, Martin RH, Zhao W, Jauch EC, Hill MD, Palesch YY. Covariate imbalance and adjustment for logistic regression analysis of clinical trial data. *Journal of Biopharmaceutical Statistics*. 2013; 23:1383–1402. [PubMed: 24138438]
19. DeMets, David. Futility approaches to interim monitoring by data monitoring committees. *Clin Trials*. 2006; 3:522–529. [PubMed: 17170036]
20. Gordan Lan KK, Wittes Janet. The b-value: A tool for monitoring data. *Biometrics*. 1988; 44:579–585. [PubMed: 3390511]
21. Gordan Lan KK, Zucker David M. Sequential monitoring of clinical trials: The role of information and brownian motion. *Statist Med*. 1993; 12:753–765.
22. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Statist Med*. 2002; 21:2917–2930.
23. Schoenfeld, David A.; Meade, Maureen O. Pro/con clinical debate: It is acceptable to stop large multicentre randomized controlled trials at interim analysis for futility. *Crit Care*. 2005; 9:34–36. [PubMed: 15693981]
24. The national heart, lung, and blood institute. ARDS clinical trials network, higher versus lower positive end-expiratory pressures in patients with acute respiratory distress syndrome. *New England Journal of Medicine*. 2004; 351:327–336. [PubMed: 15269312]

Fig. 1. CP_{Z_x} vs. Imbalance (Trial Fraction = 0.60) under the Alternative Hypothesis Assuming Power = 80%. CP_{Z_x} (using Equation (2)) under the alternative hypothesis corresponding to 80% power is depicted for various levels of covariate influence. A positive imbalance in this case suggests that the placebo group is favored at baseline with respect to the covariate of interest.

Unadjusted CP vs. Imbalance, Current Trend, $t=0.6$ **Fig. 2.**

Unadjusted CP vs. Imbalance (Trial Fraction = 0.60) under the Current Trend. Five thousand (5000) clinical trials with two treatment arms (an active treatment group and a placebo group) were simulated and unadjusted CP (using the relationship as in Equation (3)) under the current trend was calculated for each. These plots show the locally weighted least squares smoothing lines comparing CP based on current trend at trial fraction 0.60 and covariate imbalance. A positive imbalance ($t > 0$) corresponds to a favored placebo group at baseline.

Fig. 3.

Unadjusted Conditional Power Bias given Covariate Imbalance, (a) When the covariate (X) had no influence on primary outcome (Y), there was no observed association between covariate imbalance and bias in calculating unadjusted CP. (b)-(d) However, as the level of covariate influence increased, the relationship between covariate imbalance (measured by the t-statistic) and bias in unadjusted CP estimation became more evident. When the active treatment group was favored (imbalance was negative, t-statistic < 0), the unadjusted CP overestimated the appropriately adjusted CP estimates. However, when the placebo group was favored (imbalance was positive, t-statistic > 0), the unadjusted CP underestimated the appropriately adjusted CP estimates. The magnitude of bias increased as the level of covariate imbalanced increased in either direction. When imbalance was essentially zero, the bias in unadjusted CP estimates was also essentially zero.

Fig. 4.

Probability of Stopping a Clinical Trial Given Covariate Imbalance. This Figure shows the estimated probability of stopping (based on generalized linear models from the simulated data) a clinical trial at trial fraction 0.30 given current covariate imbalance for various stopping rules. Recall that a positive imbalance signifies that the placebo group was favored at baseline with respect to the covariate. The 10% Rule is equivalent to stopping the clinical trial if calculated CP is less than 10%, and so on. Unadjusted CP was calculated based on the current data trend. The scenarios shown in this Figure correspond to sample sizes of 1000 and a simulated skewed or lognormal covariate distribution. The results were consistent across sample size as well as covariate distribution, (a) The imbalance seemed to have no effect on whether a trial would be stopped for futility when there was no simulated association between covariate and outcome, (b)-(d) As the level of covariate influence increased, the amount of influence covariate imbalance had on the decision to stop the trial for futility also increased. Each plot contains a vertical line at 1.96 to serve as reference for a “statistically significant” level of imbalance (at the 5% level).

Fig. 5.

Probability of Continuing a Clinical Trial Given Covariate Imbalance. This Figure shows the estimated probability of incorrectly continuing (based on the generalized linear models from simulated data) a clinical trial at trial fraction 0.60 given current covariate imbalance for various stopping rules. In these plots, positive imbalance corresponds to favoring the active treatment group with respect to the influential covariate. The 10% Rule is equivalent to stopping the clinical trial if calculated CP is less than 10%, and so on. Unadjusted CP was calculated based on the current data trend. The scenarios shown in this Figure correspond to sample sizes of 1000 and a simulated skewed or lognormal covariate distribution. The results were consistent across sample size as well as covariate distribution, (a) Covariate imbalance seemed to have no effect on whether a trial would continue for lack of futility when there was no simulated association between covariate and outcome, (b)-(d) As the level of covariate influence increased, the effect of covariate imbalance on the decision to continue enrolling subjects also increased. Each plot contains a vertical line at 1.96 to serve as reference for a “statistically significant” level of imbalance (at the 5% level).

Table 1

Probability of Making Incorrect Trial Decisions using the 15% Stopping Rule, ρ = the correlation between covariate and outcome.

ρ	Imbalance	P(Stop), t=0.30	P(Stop), t=0.60	P(Continue), t=0.30	P(Continue), t=0.60
0	1.282	0.22	0.14	0.25	0.13
	1.645	0.22	0.14	0.25	0.13
	1.960	0.22	0.14	0.25	0.13
0.3	1.282	0.31	0.26	0.37	0.23
	1.645	0.35	0.30	0.42	0.27
	1.960	0.39	0.34	0.46	0.31
0.6	1.282	0.45	0.34	0.54	0.35
	1.645	0.57	0.45	0.65	0.48
	1.960	0.66	0.56	0.73	0.59
0.8	1.282	0.58	0.48	0.70	0.43
	1.645	0.75	0.69	0.83	0.63
	1.960	0.86	0.82	0.91	0.78