



Citation for published version:

Bartlett, JW 2018, 'Covariate adjustment and estimation of mean response in randomised trials', *Pharmaceutical Statistics*, vol. 17, no. 5, pp. 648-666. <https://doi.org/10.1002/pst.1880>

DOI:

[10.1002/pst.1880](https://doi.org/10.1002/pst.1880)

Publication date:

2018

Document Version

Peer reviewed version

[Link to publication](#)

This is the peer reviewed version of the following article: Bartlett JW. Covariate adjustment and estimation of mean response in randomised trials. *Pharmaceutical Statistics*. 2018;119, which has been published in final form at <https://doi.org/10.1002/pst.1880>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This is the peer reviewed version of the following article: 'Bartlett JW. Covariate adjustment and estimation of mean response in randomised trials. *Pharmaceutical Statistics* 2018' which has been published in final form at <https://doi.org/10.1002/pst.1880>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Covariate adjustment and estimation of mean response in randomised trials

Jonathan W. Bartlett
Statistical Innovation Group
Advanced Analytics Centre
AstraZeneca, Cambridge, UK *

Abstract

Analyses of randomised trials are often based on regression models which adjust for baseline covariates, in addition to randomised group. Based on such models, one can obtain estimates of the marginal mean outcome for the population under assignment to each treatment, by averaging the model based predictions across the empirical distribution of the baseline covariates in the trial. We identify under what conditions such estimates are consistent, and in particular show that for canonical generalised linear models, the resulting estimates are always consistent. We show that a recently proposed variance estimator underestimates the variance of the estimator around the true marginal population mean when the baseline covariates are not fixed in repeated sampling, and provide a simple adjustment to remedy this. We also describe an alternative semiparametric estimator which is consistent even when the outcome regression model used is misspecified. The different estimators are compared through simulations and application to a recently conducted trial in asthma.

1 Introduction

Analyses of randomised clinical trials often adjust for one or more baseline covariates, typically through fitting a regression model for the outcome variable, conditional on randomised group and the specified baseline covariates. The purposes of such adjustment is usually primarily to take account of chance imbalance in the distribution of the baseline covariates between treatment groups. Statistically such adjustment generally improves the power of a test of the null hypothesis of no treatment effect. For linear outcome models the estimated treatment effect is also more precise, while for non-linear outcome models the situation is more nuanced, since in general covariate adjustment alters the conditional treatment effect being targeted [1].

In addition to reporting baseline covariate adjusted estimates of treatment effect, trial reports almost always report crude or raw summary statistics such as means for the outcome, by treatment group. Since the primary treatment effect estimate is adjusted for baseline covariates, a natural although less common next step is to report estimates or predictions of mean outcome by treatment group which take the baseline covariates into account.

As described recently by Qu and Luo [2], some statistical packages provide estimates by calculating predictions for the mean outcome for a given treatment, setting the baseline covariates to their mean

*Correspondence to: Jonathan Bartlett, Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK. Email j.w.bartlett@bath.ac.uk

values in the sample - ‘prediction at the mean’. An alternative, also described by [2], involves predicting the mean outcome assuming assignment to a particular treatment for each patient in the trial, using each patient’s observed baseline covariates, and averaging these predictions across all patients in the trial. We refer to this as the ‘marginal mean’ method. For linear models, the two approaches give identical results, while for non-linear models, such as logistic or Poisson regression, in general they do not [3]. Indeed, for non-linear models, the population parameters being estimated by the two approaches differ.

The marginal mean approach estimates what the mean outcome would be if the population were assigned a given treatment. This is the same population parameter estimated by the crude treatment group specific mean outcome. In contrast, ‘prediction at the mean’ predicts the mean outcome for a patient with baseline covariate values equal to their means. As noted by others [4], when some of the covariates are categorical, such as gender or region, ‘prediction at the mean’, or prediction with other proportions assigned to the levels of categorical variables, arguably makes little sense. This is because any given patient can only belong to one level of each categorical covariate, and the resulting predicted value applies to a patient with covariate values that are never and can never be seen in the population.

Qu and Luo recently proposed using the marginal mean approach in randomised trials to estimate treatment group specific means, and advocated its use in preference over ‘prediction at the mean’ for non-linear outcome models [2]. In addition to point estimation, Qu and Luo described a variance estimator, and demonstrated that the resulting estimator was unbiased and confidence intervals achieved their nominal coverage level.

In this paper we investigate marginal mean estimation in further detail. In Section 2 we review crude marginal mean estimation in trials, whereby baseline covariates are not utilised. In Section 3 we consider the estimator described by Qu and Luo in detail. We first describe conditions under which it is consistent, and show that for certain outcome model types it remains consistent even when the outcome regression model is misspecified. We then show that when, as is typically the case, baseline covariates would not be fixed in repeated samples, the variance estimator described by Qu and Luo underestimates the true sampling variance of the estimator around the marginal (super) population mean, and we give a simple adjustment which remedies this. In Section 4 we consider an alternative semiparametric baseline adjusted estimator which is guaranteed to be unbiased irrespective of whether the outcome model is correctly specified. Next in Section 5 we consider the impact of stratified randomisation on the preceding results. In Section 7 we report simulation results comparing the different estimators. As an illustrative example in Section 8 we present results from a recently conducted trial in asthma. We conclude with a discussion in Section 9.

2 Marginal mean estimation ignoring covariates

Let Y^z denote the potential outcome for an individual patient were they, possibly contrary to fact, given treatment z [5]. We focus on inference for the (true) marginal mean outcome under treatment z , defined as

$$\mu(z) = E(Y^z) \tag{1}$$

Here the expectation is across a (super) population of patients. Consider a randomised trial which randomly samples n patients from this population and randomly assigns treatment. We let Z record the individual patient’s randomised treatment group and Y their outcome on their assigned treatment. Randomisation ensures that $\mu(z) = E(Y^z) = E(Y|Z = z)$, so that $\mu(z)$ can be estimated by the sample

mean outcome in those patients randomised to $Z = z$:

$$\hat{\mu}_1(z) = \frac{\sum_{i=1}^n 1(Z_i = z)Y_i}{\sum_{i=1}^n 1(Z_i = z)}$$

where $1(\cdot)$ denotes the indicator function which takes value one when its argument is true, and takes value zero otherwise. The mean for each treatment group is calculated and reported as a matter of course in randomised trials. Provided the trial used simple randomisation, the variance of $\hat{\mu}_1(z)$ can be estimated in the usual way for the variance of a sample mean of independent random variables.

In some trials the outcome Y represents the number of events that occur in each patient during a planned follow-up time τ . In this setting our estimand of interest may be the rate on each treatment, defined as $\mu(z) = E(Y^z)/\tau$. When all patients have complete follow-up, this can be estimated by estimating $E(Y^z)$ and dividing by τ . A complication however is that often the actual follow-up time T may be less than the planned follow-up τ for some patients, leading to censoring in Y . We defer consideration of this situation to Section 6.2.

3 Baseline adjusted marginal mean estimation

As described in the introduction, trials almost always collect baseline covariates X which are related, to a greater or lesser extent, to the outcome Y . In this section we consider estimation of $\mu(z)$ where we utilise such covariates.

3.1 Outcome model specification and estimation

We assume that the trial's analysis plan specifies a regression model for Y with X and Z as covariates. This could be a fully parametric model, or a semiparametric model, but we require that it at least specifies how the mean of Y depends on X and Z :

$$E(Y|X, Z) = h(X, Z, \beta) \tag{2}$$

for an assumed function $h(X, Z, \beta)$ and where β are the model parameters. The most common specification is that

$$h(X, Z, \beta) = g^{-1}(\beta_0 + \beta_X^T X + \beta_Z Z)$$

for some link function $g(\cdot)$. In the case of more than two treatment groups, Z would be replaced in the linear predictor by a vector of indicator functions. The outcome model may sometimes contain additional parameters η , such as the residual variance in the case of linear regression or the dispersion parameter in the case of negative binomial regression.

We let $\hat{\beta}$ denote the estimated value of β . In cases where the outcome model is fully parametric, the parameters are usually estimated by maximum likelihood. More generally, we assume that β is estimated as the value solving a set of estimating equations, which are consistent for the true value of β when the outcome model is correctly specified. We let $\tilde{\beta}$ denote the large sample probability limit of the estimator $\hat{\beta}$. This is the value that is consistently estimated as the sample size n tends to infinity.

3.2 Estimation of $\mu(z)$

To estimate $\mu(z)$ using the baseline adjusted outcome model, we can use the law of total expectation to express

$$\mu(z) = E(Y^z) = E\{E(Y^z|X)\} \quad (3)$$

This expression says that $\mu(z)$ can be calculated by predicting the mean outcome under treatment z for each patient given their baseline covariates X , and then averaging these predictions across all patients. This motivates the estimator described by Qu and Luo [2], which is given by:

$$\hat{\mu}_2(z) = n^{-1} \sum_{i=1}^n h(X_i, z, \hat{\beta}) \quad (4)$$

We emphasise that this estimator, unlike $\hat{\mu}_1(z)$, uses data from all patients, and not only those randomised to $Z = z$. It is an example of a standardisation estimator, where we are standardising to the empirical distribution of the baseline covariates of all patients in the trial [5].

We note that Qu and Luo proposed $\hat{\mu}_2(z)$ as an estimator not of $\mu(z)$, but rather of the following trial specific parameter:

$$n^{-1} \sum_{i=1}^n h(X_i, z, \beta) \quad (5)$$

This quantity is the mean of the (true) model predicted mean outcomes under treatment z for the particular patients in the trial at hand. We note that it differs in general, even if the regression model is correctly specified, from the mean response in the trial participants had they all received treatment z , which is $n^{-1} \sum Y_i^z$.

3.3 Consistency

Since $\hat{\mu}_2(z)$ is the sample mean of $h(X, z, \hat{\beta})$, it follows that $\hat{\mu}_2(z)$ is a consistent estimator of $E\{h(X, z, \tilde{\beta})\}$. It will therefore be consistent for $\mu(z)$ if and only if $E\{h(X, z, \tilde{\beta})\} = \mu(z)$. A sufficient condition for this to hold is that $E(Y|X, Z) = h(X, Z, \tilde{\beta})$. This is satisfied in particular if the outcome model is correctly specified. In general, if the outcome model is misspecified, $\hat{\mu}_2(z)$ is biased for $\mu(z)$.

There are however important cases where $E(Y|X, Z) = h(X, Z, \tilde{\beta})$ even when the outcome model is misspecified in certain respects. Specifically, when the outcome model is a generalised linear model, it is sufficient that the conditional mean $E(Y|X, Z) = h(X, Z, \beta)$ is correctly specified. For example, if the outcome model is Poisson regression, provided the conditional mean specification is correct, there is no need for Y to be Poisson distributed conditional on X and Z . A similar argument applies in the case of negative binomial regression (NB2), whereby the regression parameters are estimated (asymptotically) without bias provided the conditional mean function is correctly specified [6].

Moreover, in the special case of canonical generalised linear models (GLM), we show in Appendix A.1 that $\hat{\mu}_2(z)$ remains consistent even if the conditional mean model is misspecified. This robustness property for canonical GLMs in randomised trials was previously derived within the framework of targeted maximum likelihood by Rosenblum and van der Laan [7].

3.4 Variance

We now consider the variance of $\hat{\mu}_2(z)$, and estimators of this variance. Qu and Luo [2] proposed a variance estimator for $\hat{\mu}_2(z)$ using the delta method, in which they conditioned on the observed covariate

values $\underline{X} = \{X_i, i = 1, \dots, n\}$, $\underline{Z} = \{Z_i, i = 1, \dots, n\}$. The marginal mean estimator given by equation (4) is a non-linear transformation of $\hat{\beta}$, and so by the delta method its (conditional) variance can be estimated by

$$\widehat{\text{Var}}(\hat{\mu}_2(z)|\underline{X}, \underline{Z}) = \hat{G}_\beta \widehat{\text{Var}}(\hat{\beta}|\underline{X}, \underline{Z}) \hat{G}_\beta^T \quad (6)$$

where

$$\hat{G}_\beta = n^{-1} \sum_{i=1}^n \frac{\partial h(X_i, z, \hat{\beta})}{\partial \beta^T} \quad (7)$$

For $\widehat{\text{Var}}(\hat{\beta}|\underline{X}, \underline{Z})$, one could use the model based variance covariance matrix, or as Qu and Luo also described, a robust sandwich variance estimator [2].

Qu and Luo's variance estimator for $\hat{\mu}(z)$ is conditional on the observed covariate values, since they considered the quantity in equation (5) as the target parameter. Their variance estimator provides valid variance estimation when the target of inference is this trial specific parameter, or when the target is $\mu(z)$ if in repeated trials the covariate values would be held fixed at their observed values. Treating covariates as fixed is in agreement with conventional inference methods for regression models parameters. In the vast majority of late phase randomised trials, the individual covariate values are not fixed, in the sense that if the trial were to be repeated with a new sample of patients, the covariate values which would be observed would change.

Although not often made explicit, for inference regarding parameters of regression models, the practice of treating covariates as treated as fixed even when in truth they are not can be justified through the concept of ancillarity. Specifically, in the regression setting, the marginal distribution of the covariates is ancillary, which means inference for parameters of the conditional distribution of the outcome given the covariates can be performed as if the covariates were in fact fixed [8].

This argument does not apply however to $\hat{\mu}_2(z)$ as an estimator of $\mu(z)$, since it depends not only on the regression coefficients of the outcome model but also on the marginal distribution of the baseline covariates X . In Appendix A.2, we use estimating equation theory to derive an expression for the repeated sampling variance of $\hat{\mu}_2(z)$ as an estimator of $\mu(z)$ when X is random in repeated samples. As noted by Qu and Luo themselves, the variance estimator of equation (6) is not appropriate for this. In Appendix A.2 we show that when the outcome model is correctly specified, or is a GLM or negative binomial regression with the conditional mean function correctly specified, the Qu and Luo variance estimator is downwardly biased, as an estimator of the variance of $\hat{\mu}_2(z)$ around $\mu(z)$, by

$$n^{-1} \text{Var} \left[h(X, z, \tilde{\beta}) - \mu(z) \right]$$

The downward bias can thus be remedied by adding

$$n^{-2} \sum_{i=1}^n \{h(X, z, \hat{\beta}) - \hat{\mu}_2(z)\}^2 \quad (8)$$

to the Qu and Luo variance estimator given in equation (6). The form of this expression shows that the Qu and Luo variance estimator may be expected to be downwardly biased to a greater extent when the baseline covariates X are strongly associated with Y . Conversely, we expect the Qu and Luo variance estimator to be approximately unbiased for the variance of $\hat{\mu}_2(z)$ around $\mu(z)$ when the baseline covariates are only weakly associated with outcome.

The covariate adjusted estimator $\hat{\mu}_2(z)$ takes into account the effects on the outcome mean of any chance imbalances in the distribution of the baseline covariates X between treatment groups. One might therefore expect $\hat{\mu}_2(z)$ to give a more precise estimate of $\mu(z)$ than the unadjusted estimator $\hat{\mu}_1(z)$. In

Appendix A.3 we give a sketch proof that shows that when the outcome model is a correctly specified parametric model, $\hat{\mu}_2(z)$ is indeed more efficient than $\hat{\mu}_1(z)$.

4 Robust baseline adjusted marginal mean estimation

As described previously, the estimator $\hat{\mu}_2(z)$ can give a more precise estimate of $\mu(z)$ than $\hat{\mu}_1(z)$, by exploiting the baseline covariates and the randomisation assumption. However, it is only consistent for $\mu(z)$ in general when the outcome model used is correctly specified. In contrast, the crude mean estimator $\hat{\mu}_1(z)$ which does not use the baseline covariates, is (assuming there is no missing data) always unbiased.

As we describe in Appendix B.1, using the semiparametric theory developed by Tsiatis [9] and Zhang *et al* [10], we can construct an estimator $\hat{\mu}_3(z)$ that exploits X to improve efficiency, yet like $\hat{\mu}_1(z)$ is guaranteed to be consistent. This estimator is given by

$$\hat{\mu}_3(z) = \hat{\mu}_1(z) - n^{-1} \sum_{i=1}^n \left[\frac{1(Z_i = z) - \hat{\pi}_z}{\hat{\pi}_z} h(X_i, z, \hat{\beta}) \right] \quad (9)$$

where $\hat{\pi}_z$ denotes the sample proportion of patients randomised to $Z = z$ and $h(X, Z, \hat{\beta})$ denotes a working model prediction for $E(Y|X, Z)$. This estimator remains consistent even if the working model $E(Y|X, Z) = h(X, Z, \beta)$ is misspecified, since the second term in the expression has mean zero due to independence of X and Z . However misspecification will affect the estimator's variance, with the variance being minimized when the working model is correctly specified. In Appendix B.1 we show that the variance of $\hat{\mu}_3(z)$ can be estimated by

$$\hat{\pi}_z^{-2} n^{-2} \sum_{i=1}^n \left[1(Z_i = z) \{Y_i - \hat{\mu}_3(z)\} - \{1(Z_i = z) - \hat{\pi}_z\} \{h(X_i, z, \hat{\beta}) - \hat{\mu}_2(z)\} \right]^2 \quad (10)$$

This approach can be used for estimation of $\mu(z)$ for each treatment group. A separate model for $E(Y|X, Z = z)$ for each z could be fitted to those randomised to $Z = z$, or alternatively, a single regression model could be fitted to all treatment groups, and the model fit used to construct estimates of $E(Y|X, Z = z)$ for each value of z . Note that if one chooses $h(X, Z, \beta) = 0$, $\hat{\mu}_3(z)$ reduces to $\hat{\mu}_1(z)$.

In Section 3 we described how $\hat{\mu}_2(z)$ is consistent irrespective of whether the outcome model is correctly specified in the special case that the latter is a canonical generalised linear model. As we show in Appendix B.2, this result can also be demonstrated in this special case by showing that $\hat{\mu}_2(z) = \hat{\mu}_3(z)$, and hence $\hat{\mu}_2(z)$ inherits the robustness property of $\hat{\mu}_3(z)$.

5 Stratified randomisation

Thus far we have assumed that each patient's treatment group is assigned using simple randomisation, such that the data on each patient is independent and identically distributed. In practice, alternative randomisation schemes are often used. The most common is stratified randomisation. It is well known that such schemes introduce a dependence in the data, which if ignored, leads to conservative inferences for treatment effect estimates [11, 12, 13]. In particular, although stratified randomisation will not affect consistency of the estimators, it will affect their variance. In the following, we consider how stratified randomisation impacts on $\hat{\mu}_1(z)$, $\hat{\mu}_2(z)$, and $\hat{\mu}_3(z)$.

5.1 $\hat{\mu}_1(z)$

Stratified randomisation ensures that for the covariates used in the randomisation, their sample distribution is identical or near identical across treatment groups. This means that the distribution of these covariates in those randomised to $Z = z$ is, in expectation, closer to the population distribution than would be expected under simple randomisation. As such, we would expect the usual i.i.d. variance estimator for $\hat{\mu}_1(z)$ to be biased upwards when stratified randomisation is used. We investigate this conjecture in the simulation study described in Section 7.

5.2 $\hat{\mu}_2(z)$

We first consider $\hat{\mu}_2(z)$ when the covariates adjusted for include all of those used in the stratified randomisation. The theory developed previously for the variance of $\hat{\mu}_2(z)$ assumed that patients' data are i.i.d. In Appendix C we show that provided the outcome model is correctly specified, the variance of $\hat{\mu}_2(z)$ can be estimated as described previously in the case of simple randomisation.

Now suppose that the covariates X adjusted for are exactly those used in the stratified randomisation, i.e. that the outcome model does not adjust for any additional covariates not used in the stratified randomisation. As noted previously, under stratified randomisation, the sample distribution of X is identical, or near identical, across the treatment groups. As such, we might expect that $\hat{\mu}_2(z)$ would not have improved efficiency compared to $\hat{\mu}_1(z)$, since those randomised to treatments other than z give no additional information about the population distribution of X . To demonstrate this conjecture in a special case, suppose that there exists a single binary baseline covariate X , randomisation was performed stratified on X , and that there are two treatment groups. Suppose we fit a canonical GLM for the outcome conditional on Y and X and Z , with main effects of X and Z . The outcome model estimating equations then consist of

$$\sum_{i=1}^n \{Y_i - h(X_i, Z_i, \hat{\beta})\} \begin{pmatrix} 1 \\ X_i \\ Z_i \end{pmatrix} = 0$$

A consequence is that $\sum Z_i Y_i = \sum Z_i h(X_i, 1, \hat{\beta})$. Since the sample distribution of X is identical (or almost identical) across treatment groups as a consequence of stratified randomisation,

$$\frac{\sum_{i=1}^n Z_i h(X_i, 1, \hat{\beta})}{\sum_{i=1}^n Z_i} \approx \frac{\sum_{i=1}^n (1 - Z_i) h(X_i, 1, \hat{\beta})}{\sum_{i=1}^n (1 - Z_i)}$$

so that

$$\begin{aligned} \hat{\mu}_2(1) &= n^{-1} \sum_{i=1}^n h(X_i, 1, \hat{\beta}) \\ &\approx \frac{\sum_{i=1}^n Z_i h(X_i, 1, \hat{\beta})}{\sum_{i=1}^n Z_i} \\ &= \frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i} \\ &= \hat{\mu}_1(1) \end{aligned}$$

A similar argument naturally carries through for $z = 0$. Thus the estimators $\hat{\mu}_1(z)$ and $\hat{\mu}_2(z)$ are essentially equivalent, and there is no gain in precision over $\hat{\mu}_1(z)$.

5.3 $\hat{\mu}_3(z)$

In general, the variance estimator described assuming data are i.i.d. for $\hat{\mu}_3(z)$ is not expected to be valid under stratified randomisation. This can be seen by the fact that $\hat{\mu}_3(z)$ reduces to $\hat{\mu}_1(z)$ upon choosing $h(X, Z, \beta) = 0$, and as described previously, we expect the i.i.d. variance estimator for $\hat{\mu}_1(z)$ to be biased upwards under stratified randomisation.

6 Missing and censored data

Thus far we have assumed that baseline covariates and outcomes are fully observed. In this section we consider the impact of missingness or censoring in the outcome Y on estimation of $\mu(z)$, and comment on the handling of missingness in covariates.

6.1 Missing data

We now consider the impact of missingness in the outcome Y on validity of estimation for each of the approaches described previously. We let $R_i = 1$ indicate that the outcome Y_i is observed, while $R_i = 0$ indicates Y_i is missing.

6.1.1 $\hat{\mu}_1(z)$

The unadjusted estimator $\hat{\mu}_1(z)$ can be calculated using the complete cases (CC) in treatment group z with Y observed as

$$\hat{\mu}_1^{CC}(z) = \frac{\sum_{i=1}^n 1(Z_i = z) R_i Y_i}{\sum_{i=1}^n 1(Z_i = z) R_i}$$

This will be unbiased provided that missingness in Y is independent of Y within group z . That is, outcomes must be missing at random within treatment group z .

6.1.2 $\hat{\mu}_2(z)$

For the covariate adjusted estimator $\hat{\mu}_2(z)$, one can fit the outcome model using data from those patients with Y observed, but then, assuming X is fully observed, estimate $\mu(z)$ using the covariate values of all patients, as per equation (4). This estimator will be consistent under the weaker assumption that missingness of the outcome is independent of outcome conditional on the covariates X and Z , i.e. missing at random conditional on X and Z . This only holds in general however provided the outcome model used is correctly specified. The previously described results for robustness under misspecification for the case of canonical GLMs would only apply if missingness were at random within treatment group z .

6.1.3 $\hat{\mu}_3(z)$

Suppose again that covariates are fully observed, and the working model is fitted using those patients with outcome observed. The estimator $\hat{\mu}_3(z)$ could then be calculated as

$$\hat{\mu}_1^{CC}(z) - n^{-1} \sum_{i=1}^n \left[\frac{1(Z_i = z) - \hat{\pi}_z}{\hat{\pi}_z} h(X_i, z, \hat{\beta}) \right]$$

With full data, the estimator $\hat{\mu}_3(z)$ is consistent even when the working model is misspecified, as described in Section 4. The second term in the preceding equation will continue to have mean zero, irrespective of whether the outcome model is correctly specified, when the covariates are fully observed. However,

consistency then depends on consistency of $\hat{\mu}_1^{CC}(z)$. Thus the estimator given in the preceding equation will in general only be consistent if missingness is at random within treatment group z .

6.2 Rate estimation with censoring

In this subsection we consider the situation where the outcome is the number of events occurring in an individual through to time τ , but for some patients follow-up ends prior to τ , leading to censoring in Y . Let $Y^z(t)$ denote the number of events a given patient experiences under treatment z up to time t . Then our estimand of interest is $\mu(z) = E(Y^z(\tau))/\tau$. We let $Y(t)$ denote the number of events an individual experiences up to time t when randomised to Z . We let T denote the length of an individual patient's follow-up time when randomised to Z . The value of $Y(t)$ is thus only known for $t \leq T$.

6.2.1 $\hat{\mu}_1(z)$

The unadjusted rate estimator for treatment z is

$$\hat{\mu}_1(z) = \frac{\sum_{i=1}^n 1(Z_i = z)Y_i(T_i)}{\sum_{i=1}^n 1(Z_i = z)T_i} \quad (11)$$

This is an unbiased estimator of $\mu(z) = E(Y^z(\tau))/\tau$ provided $E(Y(T)|Z = z, T) = \mu(z)T$. This holds if the event rate is constant in time and the censoring mechanism which determines T is independent of the event process [14]. A variance estimator for this, derived in Appendix D.2, is given by

$$\hat{\tau}_z^{-2} n_z^{-2} \sum_{i=1}^n [1(Z_i = z)\{Y_i(T_i) - \hat{\mu}_1(z)\}]^2 \quad (12)$$

where n_z denotes the number of patients randomised to $Z = z$ and $\hat{\tau}_z$ denotes the mean follow-up time in those randomised to $Z = z$.

6.2.2 $\hat{\mu}_2(z)$

For the covariate adjusted estimator, typically one assumes $E(Y(t)|X, Z) = h(X, Z, \beta)t$. In statistical software this assumption is specified by passing the patient's follow-up time as an appropriate offset, for example in a Poisson or negative binomial regression. The estimator $\hat{\mu}_2(z)$ and its variance can be calculated as described previously. Estimates are consistent if the model correctly specifies $E(Y(t)|X, Z) = h(X, Z, \beta)t$ and the censoring is independent of the event process [14]. More generally, if the outcome model is a correctly specified parametric model for the event process fitted by maximum likelihood, estimates are consistent under the weaker conditionally independent censoring or censoring at random assumption [14, 15]. In the special case that Poisson regression is used for the outcome model, with $\log(T)$ as an offset, we show in Appendix D.1 that provided $E(Y(T)|Z = z, T) = \mu(z)T$ and $T \perp\!\!\!\perp X|Z = z$, $\hat{\mu}_2(z)$ is consistent for $\mu(z)$ irrespective of whether the Poisson regression is correctly specified.

6.2.3 $\hat{\mu}_3(z)$

In Appendix D.2 we show that $\hat{\mu}_3(z)$ remains as per equation (9), with $h(X, z, \beta)$ denoting the working model prediction for $E(Y(1)|X, Z = z)$. It is consistent under the same assumption required for consistency of $\hat{\mu}_1(z)$, that $E(Y(T)|Z = z, T) = \mu(z)T$. We also show in Appendix D.2 that its variance can be estimated by

$$\hat{\pi}_z^{-2} \hat{\tau}_z^{-2} n^{-2} \sum_{i=1}^n \left[1(Z_i = z)\{Y_i(T_i) - \hat{\mu}_3(z)\} - \hat{\tau}_z\{1(Z_i = z) - \hat{\pi}_z\}\{h(X_i, z, \hat{\beta}) - \hat{\mu}_2(z)\} \right]^2 \quad (13)$$

6.3 Missing covariates

We have considered the validity of simple implementations of the marginal mean estimators when some outcomes are missing or censored. The methods described here could be used in conjunction with for example multiple imputation for handling such missingness. We have also assumed throughout that the baseline covariates are fully observed. Sometimes baseline covariates contain missing values, although typically not many. In the situation when outcomes are fully observed but some baseline values are missing, certain single imputation approaches can be used [16]. With missing values in both baseline covariates and outcome, multiple imputation may be an attractive approach. Multiple imputation can also be used to perform sensitivity analyses in which deviations from the missing at random assumption are explored [17].

7 Simulations

In this section we present simulation results to investigate the finite sample properties of the methods described, first where random permuted block randomisation is used, and second where stratified randomisation is used. Treatment allocation was performed with $P(Z = 1) = 0.5$. We simulated data for trials of size $n = 400$, with a single binary covariate $X \sim \text{Bernoulli}(0.5)$. For a random 25% of patients their follow-up time T was generated from a uniform distribution on $(0, 1)$, with the remainder having follow-up $T = 1$. We investigated the estimators described previously in four different scenarios:

1. $Y_i(T_i)$ generated from a Poisson distribution, with mean $\gamma_i T_i \exp(3X_i + Z_i)$, where γ_i was independently gamma distributed with shape 2 and scale 1/2. The working model used for $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ was a negative binomial model, with main effects of X_i and Z_i , and $\log(T_i)$ as offset
2. $Y_i(T_i)$ generated in the same way, but with γ_i a log normally distributed random effect with mean 1 and variance 1/2. The working model used for $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ was again a negative binomial model with main effects of X_i and Z_i , and $\log(T_i)$ as offset
3. $Y_i(T_i)$ generated from a Poisson distribution, with mean $\gamma_i T_i \exp(3X_i + Z_i - 1.5X_i Z_i)$, where γ_i was gamma distributed with shape 2 and scale 1/2. The working model used for $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ was again a negative binomial model with main effects of X_i and Z_i , and $\log(T_i)$ as offset
4. $Y_i(T_i)$ generated from a Poisson distribution, with mean $\gamma_i T_i \exp(3X_i + Z_i - 1.5X_i Z_i)$, where γ_i was gamma distributed with shape 2 and scale 1/2. The working model used for $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ was a Poisson model with main effects of X_i and Z_i , and $\log(T_i)$ as offset

We note that these scenarios assume very strong associations between X and the outcome, since it is the strength of this association that largely drives the differences we expect to see. The variance of $\hat{\mu}_1(z)$ was estimated using equation (12). The Qu and Luo variance estimator for $\hat{\mu}_2(z)$ was calculated using equation (6), with a sandwich variance estimator used to estimate the variance of $\hat{\beta}$. The term in equation (8) was then added to this variance to allow for the fact that X was not fixed in repeated samples. The variance of $\hat{\mu}_3(z)$ was calculated using equation (13). Confidence intervals were calculated on the log scale and then back transformed. Coverage probabilities were calculated as the proportion of simulations in which the interval included $\mu(1)$.

Table 1 shows the results of the simulations with random permuted block randomisation. As expected based on Section 6.2, $\hat{\mu}_1(1)$ was unbiased for all scenarios because the rate was constant over time and the censoring was independent of the event process for all scenarios. The covariate adjusted estimator $\hat{\mu}_2(1)$ was unbiased in scenarios 1 and 2 since the conditional mean function was correctly modelled and the censoring was independent of the event process. In scenario 3, the conditional mean function in the outcome model was misspecified, and so $\hat{\mu}_2(1)$ was biased. In scenario 4 $\hat{\mu}_2(1)$ was unbiased, despite the

conditional mean function being misspecified, since a canonical GLM (Poisson) model was used and the assumptions that $E(Y(T)|Z = z, T) = \mu(z)T$ and $T \perp\!\!\!\perp X|Z = z$ were true. The robust estimator $\hat{\mu}_3(1)$ was unbiased in all scenarios since $\hat{\mu}_1(1)$ was always unbiased.

As expected, $\hat{\mu}_2(1)$ was more efficient than $\hat{\mu}_1(1)$. Confidence intervals based on equation (6) had coverage below the nominal 95% level, as predicted by theory, because the covariates were not fixed in repeated samples in the simulation. In contrast, in those scenarios where $\hat{\mu}_2(1)$ was unbiased, the coverage of the random X confidence intervals was close to the 95% level. This was the case even in scenario 4, using a misspecified Poisson model, a result which does not appear to be implied theoretically. The robust estimator $\hat{\mu}_3(1)$ was unbiased in all four scenarios, was more efficient than $\hat{\mu}_1(1)$, and confidence intervals had coverage just slightly below the 95% level.

Table 2 shows the results of simulations where randomisation was stratified on X . As conjectured, the i.i.d. confidence intervals for $\hat{\mu}_1(1)$ had coverage higher than 95%. Also, since the covariate X was stratified on in the randomisation, neither $\hat{\mu}_2(1)$ nor $\hat{\mu}_3(1)$ had improved efficiency relative to $\hat{\mu}_1(1)$. Fixed X confidence intervals for $\hat{\mu}_2(1)$ again under covered, whereas the random X intervals had correct coverage whenever $\hat{\mu}_2(1)$ was unbiased. Interestingly, confidence intervals for $\hat{\mu}_3(1)$ had approximately correct coverage despite the use of stratified randomisation, even in scenario 3 where the working conditional mean model was misspecified. Further research is warranted to understand the reason for this.

8 Illustrative analysis

In this section we re-analyse data from the Sirocco trial, a randomised placebo controlled phase 3 trial designed to investigate the efficacy and safety of benralizumab (benra) in patients with severe asthma [18]. Across 374 sites in 17 countries, 1,205 patients were randomised. Randomisation was stratified by age group, country/region, and blood eosinophil count category. The primary analysis sample consisted of those patients with blood eosinophil counts at least 300 cells per μL , in which 267 were randomised to placebo, 275 to benra 30 mg every 4 weeks, and 267 to benra 30 mg every 8 weeks. The analyses presented here are based on a de-identified version of the data with patients who revoked their informed consent removed, and which includes 257 patients on placebo, 263 patients on benra 4 week, and 256 patients on benra 8 week.

The primary outcome was asthma exacerbations during a planned 48 week follow-up period. In the subset used here, approximately 95% of patients had at least 46 of the planned 48 weeks of follow-up. The primary outcome was analysed using negative binomial regression, with randomised treatment group, region, exacerbations in the previous year (two, three, or four or more), and oral corticosteroid use at time of randomisation as covariates. As part of the trial’s analyses [18], $\hat{\mu}_2(z)$ was calculated using the negative binomial model for each of the three treatment groups, with 95% CIs calculated using the method proposed by [2].

Table 3 shows the estimated asthma exacerbation rates using $\hat{\mu}_1(z)$, $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ for each of the three treatment groups. For $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ estimates are shown both for a negative binomial and Poisson model, in which the same covariates were adjusted for as in the trial’s primary analysis. Confidence intervals are shown based on the method proposed by Qu and Luo [2], and using the adjustment given in equation (8). Sandwich variances were used to estimate the variance covariance matrix of the model regression parameters.

For the rate under placebo the covariate adjusted estimates are all smaller than the crude rate. For benra 4 weeks the crude and adjusted estimates are quite similar, while for benra 8 weeks the covariate adjusted estimates are all somewhat higher than the crude rate. Differences between the crude and adjusted estimates are to be expected since as described previously, the latter adjust the crude rates for chance imbalance in the baseline covariate distribution between randomised groups. Differences between

the various covariate adjusted estimates for each treatment were small.

The confidence intervals for $\hat{\mu}_2(z)$ constructed as proposed by Qu and Luo [2] were very slightly narrower than those that include the adjustment given in equation (8). Thus here assuming the covariates were fixed for the purposes of variance estimation of $\hat{\mu}_2(z)$ made very little difference. This can be explained by the fact that although some of the baseline covariates adjusted for were statistically significant with moderately large associations, their estimated associations were nevertheless smaller in magnitude than what was assumed in the simulation study.

9 Discussion

When a randomised trial’s analysis adjusts for baseline covariates, we believe it is advisable to also estimate and report baseline adjusted estimates of mean outcome under each treatment. Estimating the mean outcome when the baseline covariates are equal to their corresponding means arguably makes little sense when some baseline covariates are categorical. Instead, the estimator $\hat{\mu}_2(z)$ proposed by Qu and Luo [2], targets the same parameter as the unadjusted group mean $\hat{\mu}_1(z)$, but removes variation attributable to chance imbalance in the distribution of baseline covariates between randomised groups. Assuming the outcome model used is correctly specified, this leads to mean estimates with improved precision.

We have shown that the variance estimator for $\hat{\mu}_2(z)$ described by Qu and Luo [2] will in general be biased downwards for its variability around $\mu(z)$ when, as is the case usually, patients’ baseline covariate values would not be fixed in repeated sampling. This bias may however be small unless the covariate effects on outcome are large. We have proposed a simple adjustment to the Qu and Luo variance estimator, which results in a valid variance estimator for $\hat{\mu}_2(z)$ (as an estimator of $\mu(z)$) when covariates are random which is valid provided the outcome model is correctly specified.

Semiparametric methods which adjust for baseline covariates offer the opportunity for tests of treatment effect with improved power while remaining valid when the working model involved is incorrectly specified [10]. As we have shown, this methodology leads directly also to estimates of $\mu(z)$, offering improved precision relative to $\hat{\mu}_1(z)$, but unlike $\hat{\mu}_2(z)$, retaining consistency even when the working model is misspecified.

As others have described previously, use of covariate adaptive randomisation schemes such as stratified randomisation offers the potential for more precise inferences to be made. Nevertheless, this potential is only realised when the analysis correctly accounts for the randomisation scheme. As we have described, stratified randomisation causes naive i.i.d. variance estimates for $\hat{\mu}_1(z)$ to be upwardly biased. Further research is warranted to understand for which randomisation schemes the proposed variance estimators for $\hat{\mu}_2(z)$ and $\hat{\mu}_3(z)$ remain valid.

10 Acknowledgements

The author thanks Stijn Vansteelandt for helpful comments and discussion on the asymptotic theory. He also thanks the Sirocco trial study team for facilitating use of the data for the illustrative analysis.

11 Funding

This work was conducted during the author’s former employment at AstraZeneca.

12 Simulation code

R code for the simulations can be found at <https://github.com/jwb133/CovAdjMarginalMean>

References

- [1] A A Tsiatis, M Davidian, M Zhang, and X Lu. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27:4658–4677, 2008.
- [2] Y Qu and J Luo. Estimation of group means when adjusting for covariates in generalized linear models. *Pharmaceutical Statistics*, 14(1):56–62, 2015.
- [3] P W Lane and J A Nelder. Analysis of covariance and standardization as instances of prediction. *Biometrics*, 38:613–621, 1982.
- [4] C J Muller and R F MacLehose. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *International Journal of Epidemiology*, 43:962–970, 2014.
- [5] M Hernán and J Robins. *Causal Inference*. Boca Raton: Chapman & Hall/CRC, 2018.
- [6] A C Cameron and P K Trivedi. *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press, 2013.
- [7] M Rosenblum and M J van der Laan. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The International Journal of Biostatistics*, 6(1), 2010.
- [8] A Buja, R Berk, L Brown, E George, E Pitkin, M Traskin, K Zhan, and L Zhao. Models as approximations: How random predictors and model violations invalidate classical inference in regression. *arXiv*, 2014. 1404.1578v2.
- [9] A A Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006.
- [10] M Zhang, A A Tsiatis, and M Davidian. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, 64:707–715, 2008.
- [11] J Shao, X Yu, and B Zhong. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, 97:347–360, 2010.
- [12] B C Kahan and T P Morris. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine*, (4):328–340, 2012.
- [13] F A Bugni, I A Canay, and A M Shaikh. Inference under Covariate-Adaptive Randomization. Technical report, Institute for Fiscal Studies, 2016. Cemmap Working Paper CWP21/16.
- [14] R J Cook and J Lawless. *The Statistical Analysis of Recurrent Events*. New York: Springer Science & Business Media, 2007.
- [15] D F Heitjan and D B Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19:2244–2253, 1991.
- [16] I R White and S G Thompson. Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine*, 24:993–1007, 2005.
- [17] J R Carpenter and M G Kenward. *Multiple Imputation and its Application*. John Wiley & Sons, Ltd, Chichester, U.K., 2013.

- [18] E R Bleecker, J M FitzGerald, P Chanez, A Papi, S F Weinstein, P Barker, S Sproule, G Gilmartin, M Aurivillius, V Werkström, and M Goldman. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting beta 2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial. *The Lancet*, 388(10056):2115 – 2127, 2016.
- [19] W K Newey and D McFadden. Large sample estimation and hypothesis testing. In R F Engle and D L McFadden, editors, *Handbook of Econometrics*, chapter 36. Elsevier B.V., 1994.
- [20] J M Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.
- [21] A Rotnitzky and J M Robins. Analysis of semi-parametric regression models with nonignorable nonresponse. *Statistics in Medicine*, 16:81–102, 1997.

Table 1: Bias and coverage of 95% confidence intervals for $\hat{\mu}_1(1)$, $\hat{\mu}_2(1)$, $\hat{\mu}_3(1)$, across 10,000 simulations. Rel. eff. denotes the ratio of the empirical variance of $\hat{\mu}_1(1)$ to the variance of $\hat{\mu}_2(1)$ or $\hat{\mu}_3(1)$.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$\hat{\mu}_1(1)$				
Mean	3.88	3.88	2.80	2.81
95% CI Cov.	94.53	94.28	94.69	94.61
$\hat{\mu}_2(1)$				
Bias	0.00	0.00	0.18	0.00
Rel. eff.	1.28	1.28	1.14	1.22
Fixed X 95% CI Cov.	89.61	89.20	81.96	91.15
Random X 95% CI Cov.	94.41	94.20	88.87	95.08
$\hat{\mu}_3(1)$				
Bias	0.00	0.00	0.00	0.00
Rel. eff.	1.26	1.25	1.21	1.22
95% CI Cov.	94.47	94.30	94.67	94.56

Table 2: Bias and coverage of 95% confidence intervals for $\hat{\mu}_1(1)$, $\hat{\mu}_2(1)$, $\hat{\mu}_3(1)$, across 10,000 simulations, with stratified randomisation. Rel. eff. denotes the ratio of the empirical variance of $\hat{\mu}_1(1)$ to the variance of $\hat{\mu}_2(1)$ or $\hat{\mu}_3(1)$.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$\hat{\mu}_1(1)$				
Mean	3.88	3.88	2.81	2.81
95% CI Cov.	96.84	96.83	96.70	96.88
$\hat{\mu}_2(1)$				
Bias	0.00	0.00	0.18	0.00
Rel. eff.	1.02	1.02	0.94	1.00
Fixed X 95% CI Cov.	89.52	89.44	82.28	91.88
Random X 95% CI Cov.	94.19	94.39	88.65	95.21
$\hat{\mu}_3(1)$				
Bias $\hat{\mu}_3(1)$	0.00	0.00	0.00	0.00
Rel. eff.	1.00	1.00	1.00	1.00
95% CI Cov.	94.05	94.27	94.78	94.90

Table 3: Estimates of asthma exacerbation rate from the Sirocco trial with 95% confidence intervals. NB - negative binomial. ‘Fixed X’ and ‘Random X’ correspond to confidence intervals calculated assuming the covariates are fixed or random respectively.

	Placebo	Benra 4 weeks	Benra 8 weeks
$\hat{\mu}_1$	1.536 (1.287, 1.833)	0.836 (0.667, 1.049)	0.652 (0.523, 0.813)
$\hat{\mu}_2$ NB Fixed X	1.464 (1.245, 1.722)	0.819 (0.561, 1.195)	0.710 (0.456, 1.107)
$\hat{\mu}_2$ NB Random X	1.464 (1.240, 1.729)	0.819 (0.560, 1.197)	0.710 (0.455, 1.109)
$\hat{\mu}_3$ NB	1.483 (1.247, 1.764)	0.839 (0.662, 1.063)	0.676 (0.546, 0.836)
$\hat{\mu}_2$ Poisson Fixed X	1.490 (1.268, 1.751)	0.841 (0.575, 1.228)	0.676 (0.413, 1.107)
$\hat{\mu}_2$ Poisson Random X	1.490 (1.263, 1.758)	0.841 (0.574, 1.230)	0.676 (0.412, 1.108)
$\hat{\mu}_3$ Poisson	1.482 (1.247, 1.762)	0.840 (0.662, 1.065)	0.674 (0.545, 0.834)

Appendix A Asymptotic theory for $\hat{\mu}_2(z)$

A.1 Consistency

Here we show that $\hat{\mu}_2(z)$ is consistent for certain outcome models, irrespective of whether the outcome model is correctly specified. As noted in Section 3, $\hat{\mu}_2(z)$ is consistent if and only if $E\{h(X, z, \tilde{\beta})\} = \mu(z)$. Suppose that the outcome model is a canonical generalised linear model. Then if there are k treatment groups, the estimating equations for β are

$$\sum_{i=1}^n \{Y_i - h(X_i, Z_i, \beta)\} \begin{pmatrix} 1 \\ X_i \\ 1(Z_i = 1) \\ \vdots \\ 1(Z_i = k) \end{pmatrix} = 0$$

and the large sample limits $\tilde{\beta}$ satisfies

$$E \left[\{Y - h(X, Z, \tilde{\beta})\} \begin{pmatrix} 1 \\ X \\ 1(Z = 1) \\ \vdots \\ 1(Z = k) \end{pmatrix} \right] = 0$$

This means in particular that

$$\begin{aligned} 0 &= E \left[1(Z = z) \{Y - h(X, Z, \tilde{\beta})\} \right] \\ &= \pi_z \left[\mu(z) - E\{h(X, z, \tilde{\beta})\} \right] \end{aligned}$$

so that $E\{h(X, z, \tilde{\beta})\} = \mu(z)$, and hence $\hat{\mu}_2(z)$ is consistent.

A.2 Variance

In this appendix we derive an expression for the sampling variance of $\hat{\mu}_2(z)$, assuming the baseline covariates are random in repeated sampling. To find the asymptotic distribution of $\hat{\mu}_2(z)$ which accounts for X being random, we use the estimating equation theory described by Newey and McFadden [19]. In addition to the parameters β indexing the conditional mean function, the outcome model may contain additional parameters η . We let $\theta = (\beta, \eta)$ denote the combined parameter. We assume that θ is estimated by the value $\hat{\theta} = (\hat{\beta}, \hat{\eta})$ solving a set of estimating equations

$$0 = \sum_{i=1}^n m(Y_i, X_i, Z_i, \hat{\theta}) \tag{14}$$

for some estimating function $m(Y, X, Z, \theta)$. In the case where the outcome model is fitted by maximum likelihood, the estimating equations correspond to the likelihood score equations. We let $\tilde{\theta} = (\tilde{\beta}, \tilde{\eta})$ denote the large sample probability limit of the estimator.

The estimator $\hat{\mu}_2(z)$ is a so called two step estimator. Suppose that $E\{h(X, z, \tilde{\beta})\} = \mu(z)$, so that $\hat{\mu}_2(z)$ is consistent. Application of Theorem 6.1 of Newey and McFadden [19] then states that, under

regularity conditions, $\hat{\mu}_2(z)$ is asymptotically normally distributed, with mean $\mu(z)$, and variance

$$n^{-1} \text{Var} \left[h(X, z, \tilde{\beta}) - \mu(z) + G_\theta \psi(Y, X, Z) \right] \quad (15)$$

where $\psi(Y, X, Z)$ is the so called influence function of $\hat{\theta}$, and

$$\begin{aligned} G_\theta &= E \left[\frac{\partial}{\partial \theta^T} \{h(X, z, \tilde{\beta}) - \mu(z)\} \right] \\ &= \left(E \left[\frac{\partial}{\partial \beta^T} h(X, z, \tilde{\beta}) \right] \quad 0 \right) \\ &= \left(G_\beta \quad 0 \right) \end{aligned} \quad (16)$$

The influence function of $\hat{\theta}$ is equal to

$$\psi(Y_i, X_i, Z_i) = -M^{-1} m(Y_i, X_i, Z_i, \tilde{\theta})$$

where

$$M = E \left[\frac{\partial}{\partial \theta^T} m(Y, X, Z, \tilde{\theta}) \right]$$

In full generality, the variance of $\hat{\mu}_2(z)$ can be estimated by

$$\widehat{\text{Var}}\{\hat{\mu}_2(z)\} = n^{-2} \sum_{i=1}^n \left[h(X, z, \hat{\beta}) - \hat{\mu}(z) + \hat{G}_\theta \hat{\psi}(Y, X, Z) \right]^2 \quad (17)$$

where

$$\hat{\psi}(Y, X, Z) = -\hat{M}^{-1} m(Y_i, X_i, Z_i, \hat{\theta}),$$

$$\hat{M} = n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \theta^T} m(Y_i, X_i, Z_i, \hat{\theta}),$$

and

$$\hat{G}_\theta = \left(n^{-1} \sum_{i=1}^n \frac{\partial h(X_i, z, \hat{\beta})}{\partial \theta^T} \quad 0 \right)$$

In certain situations, a simplification of the preceding variance estimator is possible. The variance in equation (15) can be expanded as

$$\begin{aligned} &\text{Var} \left[h(X, z, \tilde{\beta}) - \mu(z) + G_\theta \psi(Y, X, Z) \right] \\ &= \text{Var} \left[h(X, z, \tilde{\beta}) - \mu(z) \right] + \text{Var} [G_\theta \psi(Y, X, Z)] \\ &\quad + 2\text{Cov} \left[h(X, z, \tilde{\beta}) - \mu(z), G_\theta \psi(Y, X, Z) \right] \end{aligned} \quad (18)$$

The covariance term can be expanded as

$$\begin{aligned} &\text{Cov} \left[E\{h(X, z, \tilde{\beta}) - \mu(z) | X, Z\}, E\{G_\theta \psi(Y, X, Z) | X, Z\} \right] \\ &\quad + E \left[\text{Cov}\{h(X, z, \tilde{\beta}) - \mu(z), G_\theta \psi(Y, X, Z) | X, Z\} \right] \end{aligned}$$

The second term in this expression is zero, since the first component is constant, conditional on X and

Z . Now suppose that the outcome model estimating function is conditionally unbiased, in the sense that

$$E\{m(Y, X, Z, \tilde{\theta})|X, Z\} = 0 \quad (19)$$

This condition typically holds if the outcome model is a correctly specified parametric regression model estimated by maximum likelihood, where the estimating equation for θ is the likelihood score equation (see for example page 393 of [20]). It holds for generalised linear models provided the conditional mean function is correctly specified. It is also satisfied when the outcome model is a correctly specified semiparametric conditional mean model [21]. When the condition holds, it then follows that $E\{G_\theta\psi(Y, X, Z)|X, Z\} = 0$, in which case the covariance term in (18) is zero.

Now consider the case where the outcome model is negative binomial regression, and suppose the conditional mean function is correctly specified. In this case the part of the estimating equations corresponding to β have expectation conditional on X and Z equal to zero. It also follows (see [6]) that the matrix M is block diagonal, and hence so is M^{-1} . Then we have

$$\begin{aligned} E\{G_\theta\psi(Y, X, Z)|X, Z\} &= -\begin{pmatrix} G_\beta & 0 \end{pmatrix} M^{-1} E\{m(Y, X, Z, \tilde{\theta}|X, Z)\} \\ &= -\begin{pmatrix} G_\beta & 0 \end{pmatrix} \begin{pmatrix} m_{11}^{-1} & 0 \\ 0 & m_{22}^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ \phi(X, Z) \end{pmatrix} \\ &= 0 \end{aligned}$$

Thus in this case again the covariance term in (18) is zero.

Suppose then that the covariance in equation (18) is zero. A key property of the influence function $\psi(Y, X, Z)$ is that $\hat{\theta} = (\hat{\beta}, \hat{\eta})$ is asymptotically normally distributed with mean $\tilde{\theta} = (\tilde{\beta}, \tilde{\eta})$ and variance $n^{-1}\text{Var}\{\psi(Y, X, Z)\}$. Consequently,

$$\begin{aligned} \text{Var}\{G_\theta\psi(Y, X, Z)\} &= nG_\theta\text{Var}(\hat{\theta})G_\theta^T \\ &= n\begin{pmatrix} G_\beta & 0 \end{pmatrix} \text{Var}(\hat{\theta}) \begin{pmatrix} G_\beta^T \\ 0 \end{pmatrix} \\ &= nG_\beta\text{Var}(\hat{\beta})G_\beta^T \end{aligned}$$

We then have that $\hat{\mu}_2(z)$ has variance

$$n^{-1}\text{Var}\left\{h(X, z, \tilde{\beta}) - \mu(z)\right\} + G_\beta\text{Var}(\hat{\beta})G_\beta^T \quad (20)$$

The second term in this expression corresponds to the variance estimator proposed by Qu and Luo [2]. As such, this expression shows that when the covariance term in (18) is zero, the variance estimator proposed by Qu and Luo underestimates the variance of $\hat{\mu}_2(z)$ as an estimator of $\mu(z)$ by $n^{-1}\text{Var}\left\{h(X, z, \tilde{\beta}) - \mu(z)\right\}$. Although this bias term tends to zero as n increases, it converges at the same rate as the other contribution to the variance, such that confidence intervals based on the Qu and Luo variance estimator will undercover even for large sample sizes. Equation (20) implies that a valid variance estimate for $\hat{\mu}_2(z)$ can be obtained, assuming the outcome model estimating function is conditionally unbiased, by adding

$$n^{-2} \sum_{i=1}^n \{h(X, z, \hat{\beta}) - \hat{\mu}_2(z)\}^2$$

to the Qu and Luo variance estimator given in equation (6).

A.3 Efficiency compared to $\hat{\mu}_1(z)$

In this section we give a sketch proof to show that $\hat{\mu}_2(z)$ is a more efficient estimator than $\hat{\mu}_1(z)$, when the outcome model is a correctly specified parametric model estimated by maximum likelihood. Let $f(Y|X, Z, \theta)$ denote this model, with $E(Y|X, Z) = h(X, Z, \theta)$. We assume a nonparametric model for X , and $P(Z = z|X) = \pi_z$ by randomisation. We then use semiparametric theory as described in [9] to demonstrate that $\hat{\mu}_2(z)$ is the most efficient estimator in this semiparametric model, and therefore that it is more efficient than $\hat{\mu}_1(z)$.

To derive the efficient estimator, we find the tangent space. Thus consider a parametric submodel, where $f(X|\lambda)$ is a parametric model for X . Since no model is required for $P(Z|X)$, the tangent space for this (arbitrary) parametric submodel is $\mathcal{J}_\theta \oplus \mathcal{J}_\lambda$ where

$$\mathcal{J}_\theta = \{B^{1 \times p} S_\theta(Y, X, Z) : \text{for all vectors } B^{1 \times p}\}$$

where $S_\theta(Y, X, Z)$ is the score vector corresponding to the parametric model $f(Y|X, Z, \theta)$, and

$$\mathcal{J}_\lambda = \{B^{1 \times q} S_\lambda(X) : \text{for all vectors } B^{1 \times q}\}$$

where $S_\lambda(X)$ is the score vector corresponding to the (arbitrary) parametric model for X . By definition, the tangent space for the semiparametric model is the mean square closure of all parametric submodel tangent spaces. Since θ and λ are variationally independent, the tangent space for the semiparametric model is $\mathcal{J}_\theta \oplus \mathcal{J}_1$ where

$$\mathcal{J}_1 = \{g(X) : E(g(X)) = 0\}$$

is the mean square closure of the submodel tangent spaces \mathcal{J}_λ . The estimator $\hat{\mu}_1(z)$ clearly belongs to the class of RAL estimators for this semiparametric model. The influence function for $\hat{\mu}_1(z)$ is

$$\frac{1(Z = z)}{\pi_z} \{Y - \mu(z)\}$$

The influence function of $\hat{\mu}_2(z)$ is

$$h(X, z, \theta) - \mu(z) + G_\theta \{E(S_\theta S_\theta^T)\}^{-1} S_\theta(Y, X, Z)$$

where

$$G_\theta = E \left[\frac{\partial h(X, z, \theta)}{\partial \theta^T} \right] = E \left[\frac{\partial E(Y|X, z, \theta)}{\partial \theta^T} \right]$$

We now show that $\hat{\mu}_2(z)$ is semiparametric efficient in this model. To show this, we show that the projection of the influence function of $\hat{\mu}_1(z)$ onto the tangent space is equal to the influence function of $\hat{\mu}_2$. Since the tangent space is the direct sum of two orthogonal spaces, the projection is equal to the sum of the projection onto \mathcal{J}_θ and the projection onto \mathcal{J}_1 . The projection onto \mathcal{J}_1 is equal to

$$E \left[\frac{1(Z=z)}{\pi_z} \{Y - \mu(z)\} | X \right] = h(X, z, \theta) - \mu(z)$$

The projection onto \mathcal{J}_θ is equal to

$$E \left[\frac{1(Z=z)}{\pi_z} \{Y - \mu(z)\} S_\theta^T \right] \{E(S_\theta S_\theta^T)\}^{-1} S_\theta(Y, X, Z)$$

Assuming that $E(S_\theta|X, Z) = 0$, the first expectation in this projection can be expanded as

$$\begin{aligned}
E \left[\frac{1(Z = z)}{\pi_z} \{Y - \mu(z)\} S_\theta^T \right] &= E(Y S_\theta^T | Z = z) \\
&= E \left[E(Y S_\theta^T | X, Z = z) | Z = z \right] \\
&= E \left[\int Y S_\theta^T f(Y|X, z) dY | Z = z \right] \\
&= E \left[\int Y \frac{\frac{\partial}{\partial \theta^T} f(Y|X, z, \theta)}{f(Y|X, z, \theta)} f(Y|X, z, \theta) dY | Z = z \right] \\
&= E \left[\int Y \frac{\partial}{\partial \theta^T} f(Y|X, z, \theta) dY | Z = z \right] \\
&= E \left[\frac{\partial}{\partial \theta^T} E(Y|X, z, \theta) \right] \\
&= G_\theta
\end{aligned}$$

assuming sufficient regularity conditions in order to exchange differentiation and integration. We thus conclude that $\hat{\mu}_2(z)$ is semiparametric efficient. Lastly, since $\hat{\mu}_2(z)$ is semiparametric efficient for this model, it is in particular more efficient than $\hat{\mu}_1(z)$.

Appendix B Asymptotic theory for $\hat{\mu}_3(z)$

B.1 Derivation of $\hat{\mu}_3(z)$

To apply the results of Zhang *et al* [10], we first recall that the estimator $\hat{\mu}_1(z)$ which does not exploit X can be expressed as solving estimating equations with estimating function $1(Z = z)\{Y - \mu(z)\}$. For this estimating function, application of their results then gives that all unbiased estimating functions for $\mu(z)$ may be written as

$$1(Z = z)\{Y - \mu(z)\} - \sum_{g=1}^k \{1(Z = g) - \pi_g\} a_g(X)$$

where k is the number of treatment groups and $a_g(X)$ is an arbitrary one dimensional function of X . The choice of $a_g(X)$ affects the efficiency of the resulting estimator of $\mu(z)$. The results of Zhang *et al* show that the most efficient choice is

$$a_g(X) = E[1(Z = z)\{Y - \mu(z)\} | X, Z = g]$$

which equals 0 for $g \neq z$ and equals $E(Y|X, Z = z) - \mu(z)$ for $g = z$. Thus the efficient estimator has estimating function

$$1(Z = z)\{Y - \mu(z)\} - \{1(Z = z) - \pi_z\} \{E(Y|X, Z = z) - \mu(z)\} \quad (21)$$

Since $E(Y|X, Z)$ is unknown, we may postulate and fit a parametric working model for it, $h(X, Z, \beta)$, and replace $E(Y|X, Z)$ by $h(X, Z, \hat{\beta})$ in (21). The resulting estimator of $\mu(z)$ is locally efficient, in the sense that if the working model is correctly specified, the estimator attains the semiparametric efficiency bound. Even if it is not correctly specified, the estimator remains consistent.

In a randomised trial, the randomisation probabilities π_z are known by design. Nonetheless, it turns out (P206 of [9]) that in general a more efficient estimator can be obtained by replacing the known π_z

by their empirical estimates $\hat{\pi}_z$, giving an estimating function

$$1(Z = z)\{Y - \mu(z)\} - \{1(Z = z) - \hat{\pi}_z\}\{E(Y|X, Z = z) - \mu(z)\}$$

Solving the resulting estimating equation leads to

$$\hat{\mu}_3(z) = \hat{\mu}_1(z) - n^{-1} \sum_{i=1}^n \left[\frac{1(Z_i = z) - \hat{\pi}_z}{\hat{\pi}_z} h(X_i, z, \hat{\beta}) \right]$$

The estimator $\hat{\pi}_z$ has estimating function $1(Z = z) - \pi_z$. As a so called two step estimator, it follows from Theorem 6.1 of Newey [19] that $\hat{\mu}_3$ has asymptotic variance

$$\begin{aligned} \pi_z^{-2} E \left[\left\{ 1(Z = z)\{Y - \mu(z)\} - \{1(Z = z) - \pi_z\}\{h(X, z, \tilde{\beta}) - \mu(z)\} + \{1(Z = z) - \pi_z\}\{E(h(X, z, \tilde{\beta})) - \mu(z)\} \right\}^2 \right] \\ = \pi_z^{-2} E \left[\left\{ 1(Z = z)\{Y - \mu(z)\} - \{1(Z = z) - \pi_z\}\{h(X, z, \tilde{\beta}) - E(h(X, z, \tilde{\beta}))\} \right\}^2 \right] \end{aligned}$$

It follows that the variance of $\hat{\mu}_3(z)$ can be estimated by

$$\hat{\pi}_z^{-2} n^{-2} \sum_{i=1}^n \left[1(Z_i = z)\{Y_i - \hat{\mu}_3(z)\} - \{1(Z_i = z) - \hat{\pi}_z\}\{h(X_i, z, \hat{\beta}) - \hat{\mu}_2(z)\} \right]^2$$

B.2 Canonical generalised linear models

We now show that for certain outcome model specifications, the estimator $\hat{\mu}_2(z)$ described in Section 3 actually corresponds to an estimator of the form of $\hat{\mu}_3(z)$, and as such, inherits the asymptotic properties of the latter. First, we re-express $\hat{\mu}_3(z)$ as

$$\hat{\mu}_3(z) = n^{-1} \sum_{i=1}^n \left[h(X_i, z, \hat{\beta}) + \frac{1(Z_i = z)}{\hat{\pi}_z} \{Y_i - h(X_i, z, \hat{\beta})\} \right]$$

Now suppose that the estimating equations used to estimate β include

$$\sum_{i=1}^n 1(Z_i = z) \{Y_i - h(X_i, z, \hat{\beta})\} = 0 \quad (22)$$

Then we have that $\hat{\mu}_3(z) = n^{-1} \sum_{i=1}^n h(X_i, z, \hat{\beta}) = \hat{\mu}_2(z)$. In this case, $\hat{\mu}_2(z)$ is consistent for $\mu(z)$, without requiring the outcome model to be correctly specified in any respect. Moreover, the variance estimator given in equation (10) can be used to estimate its variance, and this is valid even when the outcome model used is misspecified.

An important collection of models satisfying the condition given in equation (22) are canonical link generalised linear models (GLMs) which include an intercept and main effect of Z [6]. These include logistic regression for binary outcomes and Poisson regression for count outcomes.

Appendix C Stratified randomisation

In this appendix we consider the impact of stratified randomisation on the variance of $\hat{\mu}_2(z)$, when the covariates used in the stratified randomisation are included within the covariates X . Then a reasonable assumption is that for $i \neq j$, $Y_i \perp\!\!\!\perp (Y_j, X_j, Z_j) | (X_i, Z_i)$. That is, that conditional on a patient's covariates and treatment assignment, their outcome does not depend on the data from any other patients. The

covariance between the contributions to the estimating equation for β from any two patients is then

$$\begin{aligned} & \text{Cov} [m(Y_i, X_i, Z_i, \beta), m(Y_j, X_j, Z_j, \beta)] \\ &= \text{Cov} [E\{m(Y_i, X_i, Z_i, \beta)|X_i, Z_i\}, E\{m(Y_j, X_j, Z_j, \beta)|X_i, Z_i\}] \\ & \quad + E [\text{Cov}\{m(Y_i, X_i, Z_i, \beta), m(Y_j, X_j, Z_j, \beta)|X_i, Z_i\}] \end{aligned}$$

If the outcome model is correctly specified, $E\{m(Y_i, X_i, Z_i, \beta)|X_i, Z_i\} = 0$, and so the first covariance term will be zero. The conditional covariance in the second term is zero due to the stated conditional independence assumption. We have thus shown that under the stated conditions, the contributions to the estimating function for β from distinct patients remain uncorrelated under stratified randomisation. Conditional on β , $\hat{\mu}_2(z)$ only depends on the data through X , and since X remain i.i.d. under stratified randomisation, the variance estimator derived previously for $\hat{\mu}_2(z)$ given in equation (20) remains valid under stratified randomisation, under the stated conditions.

Appendix D Rate estimation with censoring

D.1 $\hat{\mu}_2(z)$

Consider the special case of Poisson regression with the canonical log link, with $\log(T)$ as an offset. We show that estimates of $\mu(z) = E(Y^z(\tau))/\tau$ remain consistent irrespective of whether the conditional mean function is correctly specified, under certain conditions. Following Appendix A.1, we have that

$$0 = E \left[\mathbf{1}(Z = z) \{Y(T) - h(X, Z, \tilde{\beta})T\} \right]$$

Now suppose that $E(Y(T)|Z = z, T) = \mu(z)T$, and in addition, $T \perp\!\!\!\perp X|Z = z$. Then

$$0 = \pi_z E(T|Z = z)\mu(z) - \pi_z E(T|Z = z)E\{h(X, z, \tilde{\beta})\}$$

so that $E\{h(X, z, \tilde{\beta})\} = \mu(z)$, and hence $\hat{\mu}_2(z)$ is again consistent.

D.2 $\hat{\mu}_3(z)$

The estimating function corresponding to $\hat{\mu}_1(z)$ is $\mathbf{1}(Z = z)\{Y(T) - \mu(z)T\}$. Applying the results of Zhang *et al* [10], the efficient estimating function is

$$\mathbf{1}(Z = z)\{Y(T) - \mu(z)T\} - \{\mathbf{1}(Z = z) - \pi_z\}\{E(Y(T)|X, Z = z) - \mu(z)E(T|X, Z = z)\}$$

The two conditional expectations involved are again unknown, but we can construct a feasible estimator by postulating working models for them. To motivate our choices, we make the working assumption that $T \perp\!\!\!\perp X|Z$, since if this did not hold, in general it would be the case that $E(Y(T)|Z = z, T) \neq \mu(z)T$, such that $\hat{\mu}_1(z)$ would not be consistent. Under this conditional independence assumption, $E(T|X, Z = z) = E(T|Z = z)$, and this latter quantity can be estimated by the sample mean of T in those randomised to $Z = z$, which we denote $\hat{\tau}_z$. For $E(Y(T)|X, Z = z)$ we approximate it by $\hat{\tau}_z h(X, z, \hat{\beta})$, where $h(X, Z, \beta)$ denotes a working model prediction for $E(Y(1)|X, Z)$. Again substituting $\hat{\pi}_z$ in place of π_z , this leads to

$$\begin{aligned} \hat{\mu}_3(z) &= \frac{\sum_{i=1}^n \mathbf{1}(Z_i = z)Y_i(T_i) - \{\mathbf{1}(Z_i = z) - \hat{\pi}_z\}\{h(X_i, z, \hat{\beta})\hat{\tau}_z\}}{\sum_{i=1}^n \mathbf{1}(Z_i = z)T_i} \\ &= \hat{\mu}_1(z) - n^{-1} \sum_{i=1}^n \left[\frac{\mathbf{1}(Z_i = z) - \hat{\pi}_z}{\hat{\pi}_z} h(X_i, z, \hat{\beta}) \right] \end{aligned}$$

Again using Theorem 6.1 of Newey [19], it follows after further algebra that the variance of $\hat{\mu}_3(z)$ can be estimated by

$$\hat{\tau}_z^{-2} n_z^{-2} \sum_{i=1}^n \left[1(Z_i = z) \{Y_i(T_i) - \hat{\mu}_3(z)\} - \hat{\tau}_z \{1(Z_i = z) - \hat{\pi}_z\} \{h(X_i, z, \hat{\beta}) - \hat{\mu}_2(z)\} \right]^2$$

Setting $h(X, Z, \beta) = 0$, $\hat{\mu}_3(z) = \hat{\mu}_1(z)$, and thus a variance estimator for $\hat{\mu}_1(z)$ is given by

$$\hat{\tau}_z^{-2} n_z^{-2} \sum_{i=1}^n [1(Z_i = z) \{Y_i(T_i) - \hat{\mu}_1(z)\}]^2$$