Research Article

Statistics in Medicine

Received 29 June 2014, Accepted 25 March 2015 Published online 14 April 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6508

Inconsistent treatment estimates from mis-specified logistic regression analyses of randomized trials

J. N. S. Matthews[a*†] and N. H. Badi[b]

When the difference between treatments in a clinical trial is estimated by a difference in means, then it is well known that randomization ensures unbiassed estimation, even if no account is taken of important baseline covariates. However, when the treatment effect is assessed by other summaries, for example by an odds ratio if the outcome is binary, then bias can arise if some covariates are omitted, regardless of the use of randomization for treatment allocation or the size of the trial. We present accurate closed-form approximations for this asymptotic bias when important normally distributed covariates are omitted from a logistic regression. We compare this approximation with ones in the literature and derive more convenient forms for some of these existing results. The expressions give insight into the form of the bias, which simulations show is usable for distributions other than the normal. The key result applies even when there are additional binary covariates in the model. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: asymptotic bias; baseline values; logistic regression; probit regression; randomized clinical trial

1. Introduction

Randomized trials are often analyzed using a linear or generalized linear model, so that the treatment effect can be adjusted for important baseline covariates. However, if some baseline variables cannot be measured, or if their importance is not appreciated, then they will be omitted from the model. Randomization ensures that the estimate of the treatment effect is unbiassed when relevant covariates are omitted from a linear model. This is a consequence of the unit-treatment additivity in such models [1, Chapter 5] and does not necessarily carry over to generalized linear models. Several non-linear models for which unbiassed estimators are obtained, notwithstanding the omission of covariates, are identified in [2], which also shows that the important case of binary outcomes analyzed using a logistic model is asymptotically biassed when covariates are omitted.

Numerous authors have addressed the problem of the effect of the omission of covariates in logistic regression. In biostatistical contributions, an epidemiological perspective is perhaps more common [3–8], but some authors do focus on randomized trials [2, 9–11]. Gail and colleagues [2] derive approximations for the asymptotic bias in the treatment estimator when all covariates other than the treatment indicator are omitted. The case of two general scalar covariates, one of which is fitted, and the other omitted is considered in [8, 10]. The main exposition in [10] assumes that the covariates are independent, but, as the authors explain, this restriction can be relaxed. In all these articles, Taylor series approximations are used to provide some indication of the size and direction of the bias, so the expressions derived are necessarily restricted to small parameter values, although whether it is the parameter of the fitted or omitted covariate that needs to be small varies between these contributions.

In this article, we make use of the properties of the extended skew-normal distribution [12] and an approximation of the logistic function by the probit to obtain expressions for the least false (LF)

[a] *School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, U.K.*
[b] *Statistics Department, Benghazi University, Benghazi, Libya*
**Correspondence to: J. N. S. Matthews, School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K.*
[†] *E-mail: john.matthews@ncl.ac.uk*

values [13, p.25] of the fitted covariates when other covariates are omitted. No use of Taylor series approximations is required, so the expressions give excellent numerical results for a wide range of parameter values and provide useful insight into the form the bias takes in a randomized trial. Our main result applies to a logistic regression with a single binary covariate, which we usually take to indicate the treatment allocation, and an arbitrary number of continuous covariates. The latter are assumed to follow a multivariate normal distribution, but simulation studies show that the results hold for a wider class of covariates. Explicit forms for the asymptotic bias given in [9, 10] are derived for our case and compared with that found using the skew-normal distribution. Extensions to allow additional binary covariates are possible, although these extensions require further assumptions.

In the next section, we present the expression for the LF values, and in Section 3, related work is explored. Extensions to allow additional binary covariates are discussed in Section 4, and some simulation results and numerical examples are given in Section 5. In the final section, the implications for the analysis of trials with a binary outcome are discussed.

2. Least false values

Suppose that the random variable $Y \in \{0, 1\}$ is related to a binary covariate $T \in \{-1, 1\}$ and further covariates X_1 and X_2, which have p and q dimensions respectively, by

$$\Pr(Y = 1 \mid T, X_1, X_2) = \text{expit}\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right) \tag{1}$$

where $\text{expit}(u) = \exp(u)/[1 + \exp(u)]$. If the fitted model omits X_2, that is if

$$\Pr(Y = 1 \mid T, X_1) = \text{expit}\left(\mu + \alpha T + \beta_1^T X_1\right) \tag{2}$$

is assumed to apply, then our model is mis-specified and the consequences for the maximum likelihood estimates $(\hat{\mu}, \hat{\alpha}, \hat{\beta}_1)$ are described by the theory of mis-specified models first outlined by White [14]. Briefly, as the sample size increases the maximum likelihood estimates tend not to the 'true' values, as they would for a correctly specified model, but to the LF values, $(\mu^*, \alpha^*, \beta_1^*)$. These are the values that minimize the Kullback–Liebler (KL) divergence between the fitted model and the true model. The KL divergence is $\mathsf{E}(\log[g(Y, X)/f(Y, X; \theta)])$, where $g(Y, X)$ is the true joint density of the response, Y, and covariates, X, and $f(Y, X; \theta)$ is the density under the assumed model: the expectation is taken with respect to $g(\cdot, \cdot)$. The KL divergence has much in common with a measure of distance between densities, as it is positive and vanishes only if $g(Y, X) = f(Y, X; \theta)$. In the present application, the LF values are the values of $\mu^*, \alpha^*, \beta_1^*$ such that the model (2) is as close as possible, in the KL sense, to the model in (1). A succinct treatment can be found in Chapter 2 of [13], where it is shown that the expected score statistic is zero at the LF values. This observation provides a way to obtain equations determining the LF values, which, for the present application, are as follows:

$$\mathsf{E}\left[\text{expit}\left(\mu^* + \alpha^* T + \beta_1^{*T} X_1\right)\right] = \mathsf{E}\left[\text{expit}\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right)\right] \tag{3}$$

$$\mathsf{E}\left[T\,\text{expit}\left(\mu^* + \alpha^* T + \beta_1^{*T} X_1\right)\right] = \mathsf{E}\left[T\,\text{expit}\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right)\right] \tag{4}$$

$$\mathsf{E}\left[X_{1j}\,\text{expit}\left(\mu^* + \alpha^* T + \beta_1^{*T} X_1\right)\right] = \mathsf{E}\left[X_{1j}\,\text{expit}\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right)\right], \tag{5}$$

where X_{1j} is the j-th element of $X_1, j = 1, \ldots, p$ and expectations are taken with respect to the joint distribution of (T, X_1, X_2).

We consider the case when, conditional on $T = t, X = \left(X_1^T, X_2^T\right)^T$ follows a multivariate normal distribution with mean ν_t and dispersion $\Omega, t = -1, 1$. In principle, we could allow the dispersion to change with T but analytic progress does not seem possible in this case. We also use $\nu_{t,1}, \nu_{t,2}, \Omega_{11}, \Omega_{22}, \Omega_{12}$ and Ω_{21} to denote the partition of ν_t and Ω induced by the partition of X. There is no explicit form for the expectations in (3), (4) and (5), but if we approximate the logit link with the probit, that is use $\text{expit}(u) \approx \Phi(cu)$ where $\Phi(\cdot)$ is the standard normal distribution function and $c = 16\sqrt{3}/(15\pi)$ [15, p.119], then the first two require the evaluation of integrals of the form $\int \Phi(\zeta^T u + \kappa)\phi_p(u; \omega, \Omega)du$, while the last requires $\int u_j \Phi(\zeta^T u + \kappa)\phi_p(u; \omega, \Omega)du$, where $\phi_p(\cdot; \omega, \Omega)$ denotes the density of a p-dimensional normal variable. Analytic forms are available for such integrals and perhaps are most easily found from expressions for

the density and expectation of the extended skew-normal (ESN) distribution [12]: these are reproduced in the Appendix, and further information can be found in the recent monograph by Azzalini and Capitanio [16]. Applying these results to the probit approximations to (3), (4) and (5) gives the following

$$\beta_1^* \approx \frac{\beta_1 + \Omega_{11}^{-1}\Omega_{12}\beta_2}{\sqrt{1 + c^2\beta_2^T\tilde{\Omega}\beta_2}} \tag{6}$$

$$\mu^* \approx \frac{1}{\sqrt{1 + c^2\beta_2^T\tilde{\Omega}\beta_2}}\left[\mu + \tfrac{1}{2}\beta_2^T\left\{(\nu_{1,2} + \nu_{-1,2}) - \Omega_{21}\Omega_{11}^{-1}(\nu_{1,1} + \nu_{-1,1})\right\}\right] \tag{7}$$

$$\alpha^* \approx \frac{1}{\sqrt{1 + c^2\beta_2^T\tilde{\Omega}\beta_2}}\left[\alpha + \tfrac{1}{2}\beta_2^T\left\{(\nu_{1,2} - \nu_{-1,2}) - \Omega_{21}\Omega_{11}^{-1}(\nu_{1,1} - \nu_{-1,1})\right\}\right], \tag{8}$$

where $\tilde{\Omega} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$ is the dispersion of X_2 conditional on X_1. Outline details of the derivation can be found in the Appendix. To repeat, the only approximation required for the results in (6), (7) and (8) is that of a logistic by a probit, which is well known to be highly accurate.

When T is the treatment indicator and X are baseline covariates from a randomized trial, then the assumption made earlier, namely $\mathsf{var}(X \mid T = 1) = \mathsf{var}(X \mid T = -1)$, is automatically satisfied and, additionally, $\nu_1 = \nu_{-1}$, so (8) implies that the LF value of the treatment effect α is

$$\alpha^* \approx \frac{\alpha}{\sqrt{1 + c^2\beta_2^T\tilde{\Omega}\beta_2}} = \frac{\alpha}{\tilde{q}}, \qquad \text{say.} \tag{9}$$

Apart from the degenerate cases when $\beta_2 = 0$ or when the variation in X_2 is wholly explained by that in X_1, that is $\tilde{\Omega} = 0$, (9) shows that the omission of relevant covariates means that the treatment estimator is biassed towards no effect.

3. Relation with other work

3.1. No fitted covariates other than the treatment indicator

Gail and colleagues [2] considered the bias of treatment estimates for the case when there are no fitted covariates, that is the fitted equation is simply

$$\Pr(Y = 1 \mid T) = \mathrm{expit}(\mu + \alpha T),$$

as opposed to (2), and where the omitted covariates are not restricted to being normally distributed. Finding μ^* and α^* amounts to solving Equations (3) and (4) with X_1 omitted. In [2], Taylor series expansions for small $\beta_2^T X_2$ were used to obtain the approximate solution

$$\alpha^* - \alpha \approx -\tfrac{1}{2}\beta_2^T\Omega_{22}\beta_2(\mathrm{expit}(\mu + \alpha) - \mathrm{expit}(\mu - \alpha)). \tag{10}$$

In [9, 10], a different approach was applied to the case when the true model has two scalar covariates, only one of which is included in the fitted model. As in [2], no assumption of normality was made. These authors also used a Taylor series expansion but now applied to the fitted, rather than the omitted covariate. Using the notation in the present paper, and taking T to be the fitted covariate, the approach in [10] noted that

$$\alpha^* = \tfrac{1}{2}\left[\mathrm{logit}\left(\pi_1^*\right) - \mathrm{logit}\left(\pi_{-1}^*\right)\right] = H(\alpha) \tag{11}$$

where $\pi_k^* = \mathsf{E}(\mathrm{expit}(\mu + k\alpha + \xi))$, with the expectation taken with respect to the distribution of $\xi = \beta_2^T X_2$. Strictly, it is the distribution of ξ conditional on $T = k$, but as T is a randomization indicator, this

coincides with the unconditional distribution of ξ. Expanding $H(\cdot)$ about $\alpha = 0$ [10] gives, in the case of logistic regression,

$$\alpha^* \approx \alpha H'(0) = \alpha \frac{\pi_0^* - \mathsf{E}[\text{expit}(\mu + \xi)^2]}{\pi_0^* - \pi_0^{*2}}. \tag{12}$$

As in (9), this is closer to 0 than α because $\mathsf{E}[\text{expit}(\mu + \xi)^2] > (\mathsf{E}[\text{expit}(\mu + \xi)])^2 = \pi_0^{*2}$. Exact analytic evaluation of $H'(0)$ is not possible, but further use of the approximation $\text{expit}(u) \approx \Phi(cu)$ and results due to D. B. Owen reproduced in [16, p.236] allow (12) to be written as

$$\alpha^* \approx \alpha \frac{2T(h,a)}{\Phi(h)\Phi(-h)} = \alpha \frac{T(h,a)}{T(h,1)} \tag{13}$$

where $h = c\left(\mu + \beta_2^T \nu_{,2}\right) / \sqrt{1 + c^2 \beta_2^T \Omega_{22} \beta_2}$, $a = 1/\sqrt{1 + 2c^2 \beta_2^T \Omega_{22} \beta_2}$ and $\nu_{,2}$ is the mean of X_2. $T(h,a)$ is Owen's T function [17], which is defined in the Appendix, where some pertinent properties are also described. From these, we can deduce the following: (i) α^* in (13) is always closer to 0 than α; (ii) as $|\, h \,|$ increases, α^* approaches α and (iii) the largest attenuation of α occurs when $h = 0$.

While (9) gives a bias in α that does not change with the mean of the covariates, this is not the case with (13). This is most accessibly shown by plotting, for a series of values of $\tilde{q}^{-1}$, $T(h,a)/T(h,1)$ against $P = \text{expit}\left(\mu + \beta_2^T \nu_{,2}\right)$, which is a typical response probability: for most randomized trials, P will be between 0.1 and 0.9. The figure shows that the bias correction using (9) is slightly conservative relative to (13) for most values of P.

3.2. *Covariates fitted in addition to the treatment indicator*

The approach taken in [10], unlike that in [2], can be adapted to the case when the fitted model includes covariates X_1 in addition to the treatment indicator. For any given X_1 (11) still applies, but with the expectation $\mathsf{E}\left[\text{expit}\left(\mu + k\alpha + \beta_1^T X_1 + \xi\right)\right]$ now taken with respect to the distribution of X_2 given X_1. Consequently, the bias factor $T(h,a)/T(h,1)$ still applies but with $a = 1/\sqrt{1 + 2c^2 \beta_2^T \tilde{\Omega} \beta_2}$ and

$$h = \frac{c\left[\mu + \beta_1^T X_1 + \beta_2^T \left(\nu_{,2} + \Omega_{21}\Omega_{11}^{-1}(X_1 - \nu_{,1})\right)\right]}{\sqrt{1 + c^2 \beta_2^T \tilde{\Omega} \beta_2}}. \tag{14}$$

This is of limited use because of the dependence on X_1, but replacing X_1 by its mean $\nu_{,1}$, so $h = c(\mu + \beta^T \nu)/\sqrt{1 + c^2 \beta_2^T \tilde{\Omega} \beta_2}$, provides a workable alternative that can be compared with (9) when multiple covariates are fitted.

3.3. *Probit regression*

It is widely acknowledged that in practice logistic and probit regressions can seldom be distinguished in terms of their fit to the data. As the present analyses have exploited the similarity of $\text{expit}(u)$ and $\Phi(u)$, it is natural to consider the use of probit regression as an alternative to logistic regression, that is to replace (1) and (2) with $\Pr(Y = 1 \mid T, X_1, X_2) = \Phi\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right)$ and so on. The LF values for the maximum likelihood estimators from a probit regression are essentially those in (6), (7) and (8), but with denominator $\sqrt{1 + \beta_2^T \tilde{\Omega} \beta_2}$ in place of $\sqrt{1 + c^2 \beta_2^T \tilde{\Omega} \beta_2}$, although the justification of this result is slightly different (see the Appendix for details). Consequently,

$$\alpha^* = \frac{\alpha}{\sqrt{1 + \beta_2^T \tilde{\Omega} \beta_2}} \tag{15}$$

is an exact expression for the asymptotic bias in the treatment effect that arises when some covariates are omitted from a probit regression with a treatment indicator and normal covariates.

Probit regression was also considered in [2] and [10]. The probit version of (10) is $\alpha^* \approx \alpha\left(1-\frac{1}{2}\beta_2^T\Omega_{22}\beta_2\right)$. The bias term given in [10] for $H'(0)$ for the probit case is $\mathsf{E}(\phi[\Phi^{-1}(\pi_0)])/\phi(\Phi^{-1}[\mathsf{E}(\pi_0)])$, where $\phi(\cdot)$ is the standard normal density. If the true model includes both X_1 and X_2, then when X_2 is omitted, the probit analogue of (11) applies and the bias factor can be evaluated at $\pi_0 = \pi_0^* = \Phi\left(\mu + \beta_1^T X_1 + \xi\right)$, with X_1 fixed at an arbitrary value and expectations taken over the distribution of X_2 conditional on X_1. The denominator of $H'(0)$ is $\phi(h/c)$, with h as in (14), and the numerator is $\mathsf{E}\left[\phi\left(\mu + \beta_1^T X_1 + \xi\right)\right]$. This last expectation has an analytic solution leading to

$$H'(0) = \frac{\mathsf{E}(\phi[\Phi^{-1}(\pi_0)])}{\phi(\Phi^{-1}[\mathsf{E}(\pi_0)])} = \frac{\frac{\phi(h/c)}{\sqrt{1+\mathsf{var}(\xi)}}}{\phi(h/c)} = \frac{1}{\sqrt{1+\beta_2^T\tilde{\Omega}\beta_2}}.$$

This coincides with the result from [2] for small β_2 and is the same correction factor as obtained from the use of the skew-normal distribution. The derivation in [10] assumes that α is small and our derivation of the previous expression has assumed that the covariates have a multivariate normal distribution. In all cases, the bias correction for probit regression, unlike logistic regression, depends only on the conditional variance of the omitted variables and their associated regression coefficients, and not on any measure of location.

4. Extensions of the model

The analysis presented thus far applies to a model where, apart from a binary treatment indicator, the covariates are assumed to be continuous. It is often the case that in clinical trials some baseline variables are categorical. While such variables may have more than two categories, they would usually be included in a linear predictor through dummy variables, so there is no loss in assuming that categorical covariates are binary. The values of the binary treatment indicator are assigned by randomization, so are independent of the values of the other covariates, a feature that would not be shared by a general binary covariate.

If the model in (1) were extended to include a single nontreatment binary covariate, $B \in \{-1, 1\}$, as in

$$\Pr(Y = 1 \mid T, B, X_1, X_2) = \mathrm{expit}\left(\mu + \alpha T + \gamma B + \beta_1^T X_1 + \beta_2^T X_2\right), \tag{16}$$

then the foregoing analysis of the effect of omitting X_2 from the fitted model can be adapted to this case. In this model, as in Section 2, T is a binary indicator of the randomized treatment, so is independent of B and X. Consequently, the parameters defining the distributions of B and X are unaffected by the value of T, and we take $\Pr(B = b) = \theta_b$ and $\mathsf{E}(X \mid B = b) = \nu_b, b = -1, 1$, but continue to assume that the variance is unaffected by the value of B, that is $\mathsf{var}(X \mid B = b) = \Omega$.

Under these assumptions, it follows that β_1^* is as in (6) and

$$\begin{aligned}
\alpha^* &\approx \frac{\alpha}{\sqrt{1+c^2\beta_2^T\tilde{\Omega}\beta_2}} \\
\gamma^* &\approx \frac{\gamma + \frac{1}{2}\beta_2^T\left(\left[\nu_{1,2} - \nu_{-1,2}\right] - \Omega_{21}\Omega_{11}^{-1}\left[\nu_{1,1} - \nu_{-1,1}\right]\right)}{\sqrt{1+c^2\beta_2^T\tilde{\Omega}\beta_2}} \\
\mu^* &\approx \frac{\mu + \frac{1}{2}\beta_2^T\left(\left(\nu_{1,2} + \nu_{-1,2}\right) - \Omega_{21}\Omega_{11}^{-1}\left(\nu_{1,1} + \nu_{-1,1}\right)\right)}{\sqrt{1+c^2\beta_2^T\tilde{\Omega}\beta_2}},
\end{aligned}$$

where $\nu_{b,1}, \nu_{b,2}$ is the partition of ν_b corresponding to the partition of X into X_1 and X_2. The above results are exact for probit regression, provided that the factor c^2 is omitted from the denominator.

The previous argument can be extended to several arbitrary binary covariates, $B_1, \ldots, B_K$, but only at the expense of rather restrictive assumptions about the form of $\mathsf{E}(X \mid B_1, \ldots, B_K)$. However, for the case when the binary covariates are independent of the normal covariates, the arguments developed in this article can be applied if some of the normal covariates are omitted. This would arise if there were $K-1$ dummy variables describing random allocation to $K > 2$ treatments, or if, for example binary variables

T_1 and T_2 described the main effects in a randomized trial with a 2×2 factorial treatment structure. In these cases, the omission of some normal covariates leads to the log odds ratios for the treatments being attenuated as in (9).

5. Some numerical results

5.1. Assessment of the accuracy of the approximations

The simulation results in Table I assess the accuracy of the forms of α^* for the logistic regression given in Equations (9), (10) and (13), with the last adapted as in (14) as necessary. The simulated value is found by fitting the reduced model to a sample of size 2×10^6 simulated from the full model: all calculations were performed in R , version 3.10 [18]. Three cases are presented: in the first, the true model has two normal covariates, neither of which is fitted, while in the second model, only one of these covariates is omitted. The third model has five covariates, three of which are omitted. In all cases, the normal covariates have mean 0, and unit variance and correlations are 0.5. In the first half of Table I, all $\beta_k = 0.5$, and in the remainder, all $\beta_k = 2$. The consequence of varying the size of the treatment effect is assessed by considering $\alpha = 0.5$ and $\alpha = 1.5$. It is important that the simulations correspond to realistic models, with outcome probabilities taking values that are appropriate for a clinical trial. From (1), we find that

$$\Pr(Y = 1 \mid T = \pm 1) \approx \Phi\left(\frac{c(\mu \pm \alpha + \beta^T \nu)}{\sqrt{1 + c^2 \beta^T \Omega \beta}}\right),$$

so if μ is chosen so that $\mu + \beta^T \nu = 0$, then the outcome probabilities will be around 0.5.

Table I. Values of α^* computed using simulation (sample of size 2×10^6) and the three approximations given in Equations (9), (10) and (13), for various values of the regression parameters. The normal covariates have mean 0, unit variance and pairwise correlation of $\frac{1}{2}$. The number of fitted normal covariates is p, and the number omitted is q: throughout $\mu = 0$.

	$p = 0,\ q = 2$	$p = 1,\ q = 1$	$p = 2,\ q = 3$
	$\alpha = 0.5\ \beta_k = 0.5$		
Numerical	0.433	0.482	0.308
Skew normal	0.446	0.485	0.330
Gail's method	0.408	–	–
Neuhaus *et al.*	0.434	0.481	0.309
	$\alpha = 1.5\ \beta_k = 0.5$		
Numerical	1.307	1.447	1.328
Skew normal	1.337	1.454	1.337
Gail's method	1.262	–	–
Neuhaus *et al.*	1.302	1.442	1.302
	$\alpha = 0.5\ \beta_k = 2$		
Numerical	0.206	0.347	0.227
Skew normal	0.220	0.350	0.220
Gail's method	−0.970	–	–
Neuhaus *et al.*	0.202	0.330	0.202
	$\alpha = 1.5\ \beta_k = 2$		
Numerical	0.619	1.045	0.677
Skew normal	0.661	1.051	0.661
Gail's method	−2.311	–	–
Neuhaus *et al.*	0.605	0.990	0.605

Table I shows that when $\alpha = 0.5$ and $\beta_k = 0.5$, all methods perform reasonably when no normal covariates are fitted, with that from (13) doing best. When some normal covariates are fitted, Gail's method is not applicable, but the proposed extension to (13) does well. The method based on the skew-normal approximation is conservative, as would be predicted from Figure 1 for response probabilities around 0.5. When the β_k are larger, Gail's method fails, as would be anticipated from its derivation. The method due to Neuhaus and colleagues performs better than the skew-normal factor when no normal covariates are fitted, but the skew normal does better when the fitted model contains some normal covariates. The method leading to (13) assumes that α is small, and the last two columns of Table I show that for $\alpha = 1.5$, the skew-normal approximation is again better when normal covariates are fitted and performs better relative to the method of Neuhaus *et al.* than it did for the smaller value of α.

The results in Table I apply to outcome probabilities around 0.5, as the covariates have zero mean and $\mu = 0$ throughout. Table II investigates the situation as outcome probabilities become larger, with $\mu = 2$ and $\mu = 4$ also being considered. When data are generated by (1) and a logistic model is fitted, then the left-hand columns of Table II show that α^* increases and the bias decreases as μ increases. This feature is well captured by the method of Neuhaus *et al.* and, for small β_k, by Gail's method, but is ignored by the approximation (9), which clearly does not depend on μ. If the data are generated not from (1) but from the probit version, as described in Section 3.3, and a probit regression is fitted, then matters are quite different and the simulated values of α^* are unaffected by changes in μ. The lack of dependence of (15) on μ is now appropriate, and the results in the right-hand part of Table II confirm that the skew-normal and Neuhaus methods are exact. This difference between logistic and probit regressions does not appear to be widely appreciated.

5.2. *Assessment of the effect of departures from normality*

Some simulations were carried out to assess the effect of nonnormality on the performance of the expressions for α^* in (9) and (13). Two types of departure were considered. The effect of a symmetric nonnormal distribution was assessed by generating X from a central multivariate t-distribution with four degrees of freedom, while the effect of skewness was assessed using the log-normal distribution. In the latter case, X was derived from a bivariate normal variable W with zero mean. To assess the effect of skewness in the fitted or omitted variable or both, three types of model were considered, with (X_1, X_2) taken as, respectively, $(\exp(W_1)', W_2)$, $(W_1, \exp(W_2)')$ and $(\exp(W_1)', \exp(W_2)')$, where as usual X_1 is the fitted covariate and X_2 is omitted and $'$ denotes centring to zero mean. The parameters of the t and log-normal distributions were chosen to give X_1, X_2 unit variance and correlation close to $\frac{1}{2}$, which implies that the skewness

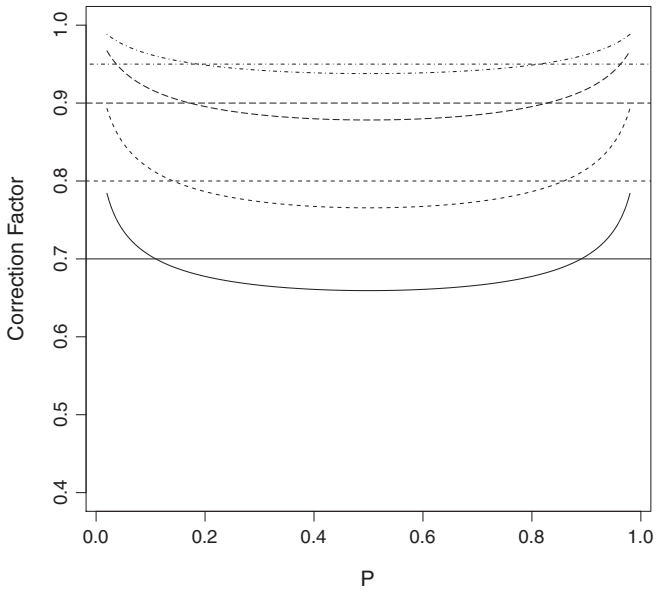

Figure 1. Correction factor $T(h, a)/T(h, 1)$ plotted against $P = \text{expit}\left(\mu + \beta_2^T \nu_{,2}\right)$, for four alternative values of the correction factor $\tilde{q}^{-1}$, namely 0.7 (solid line), 0.8 (dashed line), 0.9 (long dashed line) and 0.95 (dot-dash line). The horizontal lines are at the values of $\tilde{q}^{-1}$.

Table II. Values of α^* computed using simulation (sample of size 2×10^6) and the approximations, for both logistic and probit regression, for different locations of the linear predictor. The normal covariates have mean 0, unit variance and pairwise correlation of $\frac{1}{2}$. The number of fitted normal covariates is p, and the number omitted is q.

	Logistic regression			Probit regression		
	$\mu = 0$	$\mu = 2$	$\mu = 4$	$\mu = 0$	$\mu = 2$	$\mu = 4$
			$p = 0,\ q = 2;\ \alpha = 0.5,\ \beta_k = 0.5$			
Numerical	0.433	0.455	0.488	0.378	0.377	0.372
Skew normal	0.446	0.446	0.446	0.378	0.378	0.378
Gail's method	0.408	0.460	0.493	0.313	0.313	0.313
Neuhaus *et al.*	0.434	0.452	0.482	0.378	0.378	0.378
			$p = 0,\ q = 2;\ \alpha = 0.5,\ \beta_k = 2$			
Numerical	0.207	0.213	0.231	0.139	0.138	0.140
Skew normal	0.220	0.220	0.220	0.139	0.139	0.139
Gail's method	−0.971	−0.139	0.390	−2.50	−2.50	−2.50
Neuhaus *et al.*	0.202	0.208	0.227	0.139	0.139	0.139
			$p = 2,\ q = 3;\ \alpha = 0.5,\ \beta_k = 0.5$			
Numerical	0.437	0.445	0.458	0.378	0.378	0.379
Skew normal	0.446	0.446	0.446	0.378	0.378	0.378
Neuhaus *et al.*	0.434	0.452	0.482	0.378	0.378	0.378

Table III. Values of α^* computed using simulation (sample of size 2×10^6) and the two approximations given in Equations (9) and (13), for various values of the regression parameters. The covariates have a multivariate t distribution with four df or are a mixture of normal and log-normal variables. In each case, one covariate is fitted and one omitted, in addition to the treatment indicator: throughout $\mu = 0$. The approximations below the line apply to all the cases above it.

	$\alpha = 0.5,\ \beta_k = 0.5$	$\alpha = 1.5,\ \beta_k = 0.5$	$\alpha = 0.5,\ \beta_k = 2$	$\alpha = 1.5,\ \beta_k = 2$
X bivariate t, four df	0.484	1.456	0.376	1.129
$X = (\exp(W_1), W_2)$	0.479	1.441	0.352	1.061
$X = (W_1, \exp(W_2))$	0.488	1.460	0.403	1.194
$X = (\exp(W_1), \exp(W_2))$	0.481	1.452	0.375	1.131
Skew-Normal	0.485	1.454	0.350	1.051
Neuhaus *et al.*	0.481	1.442	0.330	0.990

df, degrees of freedom.

of the log-normal variables is 2.84. In all simulations $\mu = 0$, with $\beta_k = 0.5$ or 2 and $\alpha = 0.5$ or 1.5, and one scalar covariate is fitted and one omitted. The correction factors $\tilde{q}^{-1}$ and $T(h, a)/T(h, 1)$ both depend solely on the mean and dispersion of the X_i's, so these will be the same for all of the models.

From Table III, we see that for smaller β_k, the predictions of bias provided by (9) and (13) remain accurate even when the covariates have nonnormal distributions. For larger values of β_k, α^* tends to be closer to α for these nonnormal covariates than for normal covariates. However, it should be noted that in this context, $\beta_k = 2$ is a large coefficient for a covariate with unit variance and will seldom be encountered in practice.

5.3. *Examples: the SNAP trial and the Mayo Clinic PBC trial*

No direct evaluation of the previous results is possible as they are all expressed in terms of parameters. However, some practical indication of the size of the asymptotic bias, and how this changes with the included covariates, can be provided by substituting estimates for the parameters from relevant studies.

The Scottish and Newcastle antiemetic pretreatment for paracetamol poisoning study (SNAP) [19] trial was designed to assess ways to reduce adverse effects in the treatment of paracetamol poisoning.

The trial used a 2×2 design, comparing (i) the standard versus a modified *N*-acetylcysteine regimen and (ii) pretreatment with an antiemetic (ondansetron) or placebo. The primary outcome was binary, namely whether or not the patient retched or vomited within 2 h of the initiation of *N*-acetylcysteine therapy. In the trial report [20], odds ratios adjusted for stratification variables were reported. The stratification variables were all binary, assessing the timing and amount of paracetamol ingestion and risk factors for hepatotoxicity. For illustration of the methods used in this paper, the data have been re-analyzed, replacing the stratification variables by related continuous covariates. These are the concentration of paracetamol in the blood at presentation (mg/L) and two measures of liver function, namely gamma glutamyl transferase (GGT) and alanine aminotransferase, both in IU/L: the liver enzyme concentrations were logged before analysis.

The dispersion matrix of the three baseline covariates, based on the 217 patients in the trial, was used as Ω and β was taken to be the estimated regression coefficients from the full logistic regression. The correlations, standard deviations and estimated regression coefficients are shown in Table IV.

The values of inflation factors, that is the amount by which α exceeds α^*, namely $\tilde{q}$ and $T(h,1)/T(h,a)$, are shown in the left-hand part of Table V for a sequence of fitted models. Initially, only the treatment indicators are fitted, and the paracetamol concentration and liver function enzymes are omitted.

The first row of Table V shows that the treatment effects had all three of these covariates been included, the so-called full conditional effects, are 13% larger than the marginal effects. The full conditional effects are about 7% larger than those that would be found from a model that included just log GGT in addition to the treatment effects. Also, adding paracetamol to the model reduces the discrepancy to around 3%. The order in which covariates are added matters - the full conditional effects are around 9% larger than those from a model with just paracetamol. These figures relate to (13), and those from (9) are slightly smaller.

An example in which the difference between marginal and conditional estimates is even more pronounced is the trial comparing penicillamine with placebo for the treatment of primary biliary cirrhosis (PBC) [21], with the data given in [22]. While the primary endpoint is a survival time, an illustration of the methods in the current paper is provided by a secondary analysis of end-of-study mortality. This illustrative analysis considers a full model that includes a treatment indicator and the logarithms of the baseline concentrations of bilirubin (mg/dL), alkaline phosphatase (IU/L) and urinary copper excretion (μg/day). It is shown in the right-hand part of Table V that the treatment effect, estimated from a model with all three covariates, is around 30-35% larger than the marginal estimate. However, once account is taken of bilirubin, the other two covariates have little additional effect. If either of the covariates other than bilirubin were fitted first, the change in the inflation factor is far less marked (to 1.216 with copper and to 1.142 for alkaline phosphatase).

Table IV. The correlations obtained from the dispersion matrix for the three continuous covariates chosen from the SNAP trial, with standard deviations on the diagonal and regression coefficients in the second column.

	β_k	Paracetamol conc.	Log GGT	Log ALT
Paracetamol conc.	0.0054	85.34		
Log GGT	−0.873	−0.093	0.812	
Log ALT	0.660	−0.150	0.448	0.686

SNAP, Scottish and Newcastle antiemetic pretreatment for paracetamol poisoning study; GGT, gamma glutamyl transferase; ALT, alanine aminotransferase.

Table V. The values of $\tilde{q}$ (cf. (9)) and $T(h,1)/T(h,a)$ (cf. (13)) for a series of increasing models for both the SNAP and PBC trials.

SNAP Trial			PBC Trial		
Included covariates	$\tilde{q}$	$T(h,1)/T(h,a)$	Included covariates	$\tilde{q}$	$T(h,1)/T(h,a)$
None	1.104	1.129	None	1.289	1.346
+ Log GGT	1.056	1.070	+ Log bilirubin	1.034	1.042
+ Paracetamol	1.028	1.035	+ Log alkaline phosphatase	1.012	1.014

SNAP, Scottish and Newcastle antiemetic pretreatment for paracetamol poisoning study.

6. Discussion

One of the most widely cited instances of the effect of omitting a covariate is Simpson's paradox [23], in which the effect of a binary covariate can be reversed when a second binary covariate is taken into account. This phenomenon was thoroughly investigated and set in more general contexts by Samuels [24]. The notion of association reversal (AR) was introduced, and the relation between AR and the amalgamation paradox (AMP) defined by Good and Mittal [25], where conditioning on a second covariate significantly alters the relationship between the outcome and the first covariate, was discussed. Samuels shows that, in general, omitting a covariate from a logistic regression can lead to AR. In our application, this would amount to α and α^*, and/or β_1^* and β_1 having opposite signs. Our analysis, as shown by (6), (7) and (8) confirms this. It also confirms Samuels's observation that if the true coefficient of the omitted variable vanishes, that is $\beta_2 = 0$, then AR cannot occur and the AMP does not apply. For $\beta_2 \neq 0$, our equations also confirm, and quantify, the result that if the fitted and omitted covariates are independent, then AR cannot occur but that the AMP is inevitable. Independence of the fitted and omitted covariates implies $\nu_{1,2} = \nu_{-1,2}$ and $\Omega_{12} = 0$, so the signs of α and α^*, and of β_1 and β_1^*, must coincide. However, both will be shrunk towards zero because $\tilde{\Omega}$ does not vanish when $\Omega_{12} = 0$. As Samuels points out, this is in contrast to the situation for linear regression, where independence of fitted and omitted covariates prevents the AMP (and hence AR). These properties are related to the non-collapsibility of the odds ratio, as discussed in [26, 27].

A consequence of this observation is that if a covariate, T, is independent of all other covariates in the model, then in a linear regression, the expectation of the estimator of its coefficient, $\mathsf{E}(\hat{\alpha})$, will be unaffected whatever other covariates are included. This is of fundamental importance for estimation of the treatment effect in a randomized trial, where the act of randomization ensures that the treatment indicator T is independent of the other baseline covariates. The reasons for taking account of baseline covariates are summarized in a review of covariate adjustment [28], and the first advantage adduced is that an adjusted analysis can correct for imbalance between treatment groups in prognostic covariates that arise despite randomization. Although in a given trial adjusted and unadjusted analyses can produce different values for $\hat{\alpha}$, they ultimately estimate the same quantity because the expectation of $\hat{\alpha}$ over the joint distribution of the covariates and response is always α.

As has been widely recognized [11, 29–33], this situation does not carry over to randomized trials with a binary outcome when logistic regression is used to adjust for baseline covariates. In this case the AMP obtains, and the estimator of the treatment effect, $\hat{\alpha}$, has an expectation that depends on which baseline variables are included and which are excluded from the analysis, so there is no longer a single, unambiguous treatment effect. The *marginal* treatment odds ratio ignores baseline covariates, and is found from fitting a logistic model which includes only T and is therefore averaged over all the baseline covariates that affect the outcome: it is also referred to as the *population averaged* effect and has the advantage of being unambiguously defined. However, one of the advantages of including appropriate baseline covariates adduced by Yu and colleagues [28] is that it provides a conditional treatment estimate that is clinically more relevant because, by taking account of the different characteristics of patients, it gives a more pertinent comparison (see [11] and Senn in discussion of [34]). However, when obtained by logistic regression, it is not uniquely defined, which is something triallists could find unsettling as the aim of a clinical trial is often thought to be to estimate *the* treatment effect.

Indeed, the multiplicity of possible treatment effects has implications for nomenclature. The present paper has used the term asymptotic bias, occasionally shortened to bias, to refer to the differences between the parameters that purport to measure the treatment effect. Although this terminology is in keeping with the other contributions to the field, it is rather misleading because it implies that there is a single true effect, with respect to which the others are biassed. The marginal treatment odds ratio is shrunk towards zero relative to the conditional ones, as shown by (9), as indeed are any conditional estimates relative to that based on a superset of the covariates. As such, the term 'attenuation', as previously suggested [11], is probably preferable to bias.

While the qualitative effect on the treatment estimator of including more baseline variables in a logistic regression has been appreciated for some time, the current paper provides a quantitative assessment of the effect. When planning a trial, either (9) or (13) will provide the triallist with some indication of the potential effect on the treatment estimate of including various sets of covariates in a logistic regression. Strictly, the attenuation factors in (9) and (13) assume that the covariates are normally distributed, but as shown in Section 4 and Section 5.2, the formulae are useful more widely. Table V shows that successively including log GGT and paracetamol each result in similar changes in the treatment effect and that it may

be useful to ensure that both are included in an analysis. On the other hand, in the PBC trial, once the effect of bilirubin has been taken into account, there is little to be gained by including further variables. In general, if the triallist can *a priori* identify a set, S, of covariates such that further additions to the set cause only small changes in $\tilde{q}$, then if variables in S are always included in the analysis, the range of conditional treatment estimators that might arise from further modification to the model may not differ to any material extent.

How baseline covariates should be selected for trials has been the subject of much discussion, for example [32, 35–38], and the general conclusion is that ideally the selection should be made *a priori*. While it is conceded that this may not always be practical [29], it would be wise to try to make the assessment of which covariates to include on the basis of pilot data rather than data from the trial itself. Given a set of covariates, the investigator can use (9) to assess how the conditional treatment effect changes with which of these covariates is selected. However, conclusions drawn on this basis can be undermined by the existence of an unknown covariate, which has an important effect on the outcome and is not closely related to the known covariates. This would not be a serious issue for trials analyzed using a linear regression because randomization provides protection against the untoward influence of unknown covariates. However, this benefit of randomization may be less valuable for binary outcomes analyzed using a logistic regression.

A further reason for recommending the use of baseline covariates [28, 32, 39] is to increase the power of an analysis, because taking account of highly prognostic covariates will markedly reduce the residual variance. Variables that have had a role in the stratification of the treatment allocation have a special status because these variables need to be included if a correct estimate of variance is to be obtained [32, 38] [40, pp. 601-2]. The scope of such recommendations is often not precisely specified, although the arguments behind them are usually rehearsed in terms of a linear regression. However, the reduction in the standard error of the treatment estimator, which occurs for a linear regression, is known not to apply to logistic regression [5, 11]. Nevertheless, including prognostic baseline covariates in a logistic regression does tend to increase the power of the study [41, 42], presumably because the increase in standard error due to the covariates is slight compared with the inflation of the treatment effect implied by (9). However, both these articles based their conclusions on simulation studies in which a single covariate was considered and the increase in power when the covariate was included was most noticeable if it was highly prognostic, so that the attenuation in (9) was marked. Whether including further covariates would lead to an increase in power is moot, because the consequent inflation in the treatment effect may no longer outweigh the increase in standard error. It should also be remarked that the rationale for the automatic inclusion of stratification variables in a logistic regression is less apparent than it is for a linear regression and might usefully be investigated further.

In practice, an investigator will often plan a trial on the basis of the power of an unadjusted analysis, although even here a realistic value for the odds ratio under the alternative may need to be judged in the light of the preceding discussion. The inclusion of an important baseline covariate, or covariates, may increase the power of the study, but the size of the effects reported in [41, 42] will not make the use of unadjusted power unreasonably conservative unless some of the covariates are very highly associated with the outcome. If the triallist has adequate information on the relationship between covariates and outcome, then more sophisticated methods based on a postulated logistic model can be used [43]: this methodology requires the evaluation of the information matrix at the alternative hypothesis. The approach in [43] uses an ingenious conditioning argument so that only one-dimensional numerical integration is required. It is possible that the approach in the present paper, using a probit approximation and the ESN distribution, could be used to go further and replace the numerical integration with an accurate analytic approximation.

The comparison between logistic and probit analyses is interesting. If the parameter estimates from a logistic regression are $\hat{\beta}$, then the estimates obtained from fitting a probit regression to the same data will be approximately $c\hat{\beta}$, so the corrections in (9) and (15) are essentially equal. However, if P is not close to 1 or 0, Figure 1 shows that the asymptotic bias α^* can be greater than is implied by (9). However, the correction in (15) is an exact result, so it may be that the problem of asymptotic bias in the estimates of the treatment effect is less if probit is preferred to logistic regression. A deeper aspect of the analysis, which may warrant further study, is the extent to which trials with a binary outcome are best served by either a logit or probit link. Both links provide protection against estimated probabilities outside [0,1], and sufficiency arguments favour a logistic link, but both bring the problems of interpretation of the treatment effect discussed previously. Gail and colleagues [2] pointed out the superior bias properties of

identity or log links, and perhaps methods of incorporating baseline covariates into analyses of absolute differences of response probabilities should be given further consideration.

Appendix

A.1. The extended skew-normal distribution

The density of an extended multivariate skew-normal (ESN) random variable $U \in \mathbb{R}^p$ [12] is

$$f(u) = \frac{\phi_p(u;\omega,\Omega)\Phi(\zeta^T(u-\omega)+\psi)}{\Phi(\psi/\sqrt{1+\zeta^T\Omega\zeta})}, \tag{A.1}$$

where ζ is a p-dimensional parameter, ψ is a scalar, $\phi_p(\cdot;\omega,\Omega)$ is the p-dimensional multivariate normal density with mean ω and dispersion Ω and $\Phi(\cdot)$ is the standard normal distribution function. The mean of the ESN distribution is

$$\mathsf{E}(U) = \omega + \frac{\Omega\zeta}{\sqrt{1+\zeta^T\Omega\zeta}}\frac{\phi(\bar{\psi})}{\Phi(\bar{\psi})},$$

where $\bar{\psi} = \psi(1+\zeta^T\Omega\zeta)^{-\frac{1}{2}}$ and $\phi(\cdot) = \phi_1(\cdot;0,1)$. From $\int f(u)du = 1$, we see that $\int \phi_p(u;\omega,\Omega)\Phi(\zeta^T(u-\omega)+\psi)du = \Phi(\psi/\sqrt{1+\zeta^T\Omega\zeta})$, with a similar manipulation giving

$$\int u_j\phi_p(u;\omega,\Omega)\Phi(\zeta^T(u-\omega)+\psi)du = \omega_j\Phi(\bar{\psi}) + \frac{(\Omega\zeta)_j}{\sqrt{1+\zeta^T\Omega\zeta}}\phi(\bar{\psi}).$$

A.2. Least false values for logistic regression

Writing $p_t = \Pr(T=t)$ and applying the approximation $\text{expit}(u) \approx \Phi(cu)$ to (3), (4) and (5) and using the properties of the ESN distribution, we obtain from (3) and (4) the equations

$$p_1\Phi\left(\psi_1^*\right) \pm p_{-1}\Phi\left(\psi_{-1}^*\right) = p_1\Phi(\psi_1) \pm p_{-1}\Phi(\psi_{-1}),, \tag{A.2}$$

and from (5), we obtain

$$\begin{aligned} &p_1\left[\nu_1\Phi\left(\psi_1^*\right) + \frac{c\Omega_{11}\beta_1^*}{\sqrt{1+c^2\beta_1^{*T}\Omega_{11}\beta_1^*}}\phi\left(\psi_1^*\right)\right] + p_{-1}\left[\nu_{-1}\Phi\left(\psi_{-1}^*\right) + \frac{c\Omega_{11}\beta_1^*}{\sqrt{1+c^2\beta_1^{*T}\Omega_{11}\beta_1^*}}\phi\left(\psi_{-1}^*\right)\right] \\ &= p_1\left[\nu_1\Phi(\psi_1) + \frac{c(\Omega\beta)_1}{\sqrt{1+c^2\beta^T\Omega\beta}}\phi(\psi_1)\right] + p_{-1}\left[\nu_{-1}\Phi(\psi_{-1}) + \frac{c(\Omega\beta)_1}{\sqrt{1+c^2\beta^T\Omega\beta}}\phi(\psi_{-1})\right] \end{aligned} \tag{A.3}$$

Here, $\beta^T = \left(\beta_1^T, \beta_2^T\right)^T$, $(\Omega\beta)_1$ denotes the first p elements of $\Omega\beta$ and

$$\psi_1^* = \frac{c\left(\mu_1^*+\alpha^*\right)}{\sqrt{1+c^2\beta_1^{*T}\Omega_{11}\beta_1^*}} \qquad \psi_{-1}^* = \frac{c\left(\mu_{-1}^*-\alpha^*\right)}{\sqrt{1+c^2\beta_1^{*T}\Omega_{11}\beta_1^*}}$$
$$\psi_1 = \frac{c(\mu_1+\alpha)}{\sqrt{1+c^2\beta^T\Omega\beta}} \qquad \psi_{-1} = \frac{c(\mu_{-1}-\alpha)}{\sqrt{1+c^2\beta^T\Omega\beta}}$$

with $\mu_t^* = \mu^* + \beta_1^{*T}\nu_{t,1}$ and $\mu_t = \mu + \beta^T\nu_t$, where $\nu_{t,1}$ is written for the first p elements of ν_t. From (A.2), we obtain $\psi_1^* = \psi_1$ and $\psi_{-1}^* = \psi_{-1}$, and using this in (A.3), we obtain

$$\frac{\Omega_{11}\beta_1^*}{\sqrt{1+c^2\beta_1^{*T}\Omega_{11}\beta_1^*}} = \frac{(\Omega\beta)_1}{\sqrt{1+c^2\beta^T\Omega\beta}} = \frac{\Omega_{11}\beta_1+\Omega_{12}\beta_2}{\sqrt{1+c^2\beta^T\Omega\beta}},$$

and these can be solved to give (6), (7) and (8).

A.3. Owen's T function

Owen's T function, which appears in (13), is defined as

$$T(h,a) = \frac{1}{2\pi}\int_0^a \frac{\exp\left[-\frac{1}{2}h^2(1+x^2)\right]}{1+x^2}dx,$$

and has an important role in the computation of bivariate normal probabilities. It can be evaluated conveniently by the function `T.Owen` in the R package `sn` [44]. For fixed $h, T(h,a)$ is an increasing function of its second argument, and as in the present application $0 < a < 1$, it follows that the expression for α^* in (13) is always closer to 0 than α. For fixed $a, T(h,a)/T(h,1)$ is an even function of h and increases as the magnitude of h increases, so the largest attenuation of α occurs at $h = 0$. As the magnitude of h increases, $T(h,a)/T(h,1)$ approaches one, so α^* approaches α.

A.4. Least false values for probit regression

The LF equations for the maximum likelihood estimators for a probit regression differ from (3) to (5) because of the presence of a weighting factor $\omega = \omega(T, X_1) = \omega(\eta^*)$ with $\eta^* = \mu^* + \alpha^* T + \beta_1^{*T} X_1$, that is the $p+2$ equations are

$$\mathsf{E}[\omega(\eta^*)Z\Phi(\eta^*)] = \mathsf{E}\left[\omega(\eta^*)Z\Phi\left(\mu + \alpha T + \beta_1^T X_1 + \beta_2^T X_2\right)\right] \tag{A.4}$$

where $\omega(\eta^*) = \phi(\eta^*)/[\Phi(\eta^*)\Phi(-\eta^*)]$, and where Z is taken to be, successively, $1, T$ and X_{1j}, $j = 1, \ldots, p$. The presence of ω means that the skew-normal distribution cannot be used to evaluate the expectations in the way it was used for logistic regression, but it can be applied to evaluate the right-hand expectation in (A.4) over the distribution of X_2 conditional on T and X_1, giving

$$\mathsf{E}\left[\omega(\eta^*)Z\Phi\left\{\frac{\mu + \beta_2^T\left(\nu_{T,2} - \Omega_{21}\Omega_{11}^{-1}\nu_{T,1}\right) + \alpha T + \left(\beta_1 + \Omega_{11}^{-1}\Omega_{12}\beta_2\right)^T X_1}{\sqrt{1+\beta_2^T\tilde{\Omega}\beta_2}}\right\}\right].$$

Consequently, if we choose β_1^*, μ^* and α^* as in (6), (7) and (8) but with denominator $\sqrt{1+\beta_2^T\tilde{\Omega}\beta_2}$ as opposed to $\sqrt{1+c^2\beta_2^T\tilde{\Omega}\beta_2}$, then Equations (A.4) will be satisfied.

Acknowledgements

The authors are very grateful to Professor D. N. Bateman and Dr S. C. Lewis of the University of Edinburgh for allowing access to the data from the SNAP trial. The authors also thank the referees, associate editor and joint editor for their helpful comments and additional references.

References

1. Cox DR. *Planning of Experiments* Wiley Classics Library edn. Wiley: Chichester, 1992.
2. 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.
3. Lee LF. Specification error in multinomial logit models. *Journal of Econometrics* 1982; **20**:197–209.
4. Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. *Journal of Clinical Epidemiology* 1991; **44**:77–81.
5. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 1991; **59**:227–240.
6. Begg MD, Lagakos S. Loss in efficiency caused by omitting covariates and misspecifying exposure in logistic regression models. *Journal of the American Statistical Association* 1993; **88**:166–170.
7. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**:948–963.
8. Drake C, McQuarrie A. A note on the bias due to omitted confounders. *Biometrika* 1995; **82**:633–638.
9. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 1991; **59**:25–35.

10. Neuhaus JM, Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika* 1993; **80**:807–815.
11. 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.
12. Arnold BC, Beaver RJ. Hidden truncation models. *Sankhyā A* 2000; **62**:23–35.
13. Claeskens G, Hjort NL. *Model Selection and Model Averaging*, IMS Monographs. Cambridge University Press: Cambridge, 2008.
14. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; **50**:1–25.
15. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions* 2nd edn, Vol. 2. Wiley: Chichester, 1995.
16. Azzalini A, Capitanio A. *The Skew-Normal and Related Families*. Cambridge University Press: Cambridge, 2014.
17. Owen DB. Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics* 1956; **27**:1075–1090.
18. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria, 2014. Available from: http://www.R-project.org.
19. Thanacoody KHR, Gray A, Dear JW, Coyle J, Sandilands EA, Webb DJ, Lewis S, Eddleston M, Thomas SHL, Bateman DN. Scottish and Newcastle antiemetic pre-treatment for paracetamol poisoning study (SNAP). *BMC Pharmacology and Toxicology* 2013; **14**:20.
20. Bateman DN, Dear JW, Thanacoody HKR, Thomas SHL, Eddleston M, Sandilands EA, Coyle J, Cooper JG, Rodriguez A, Butcher I, Lewis SC, Vliegenthart ADB, Veiraiah A, Webb DJ, Gray A. Reduction of adverse effects from intravenous acetylcysteine treatment for paracetamol poisoning: a randomised controlled trial. *Lancet* 2014; **383**:697–704.
21. Dickson ER, Fleming TR, Wiesner RH, Baldus WP, Fleming CR, Ludwig J, McCall JT. Trial of penicillamine in advanced primary biliary cirrhosis. *New England Journal of Medicine* 1985; **312**:1011–1015.
22. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis* 2nd edn. Wiley: Chichester, 2005.
23. Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B* 1951; **13**:238–241.
24. Samuels ML. Simpson's paradox and related phenomena. *Journal of the American Statistical Association* 1993; **88**:81–88.
25. Good IJ, Mittal Y. The amalgamation and geometry of two-by-two contingency tables. *Annals of Statistics* 1987; **15**: 694–711.
26. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science* 1999; **14**:29–46.
27. Clogg CC, Petkova E, Shihadeh ES. Statistical methods for analyzing collapsibility in regression models. *Journal of Educational Statistics* 1992; **17**:51–74.
28. Yu L, Chen A, Hopewell S, Deeks JJ, Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials* 2010; **11**:59.
29. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 2002; **21**:2917–2930.
30. Martens EP, Pestman WR, Klungel OH. Letter to the editor. *Statistics in Medicine* 2007; **26**:3205–3212.
31. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**:1049–1060.
32. Altman DG. Covariate imbalance, adjustment for. In *Encyclopedia of Biostatistics*, 2nd edn, Armitage P, Colton T (eds). Wiley: Chichester, 2005; 1273–1278.
33. 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.
34. Lee Y, Nelder JA. Conditional and marginal models: another view. *Statistical Science* 2004; **10**:219–238.
35. Armitage P, Gehan EA. Statistical methods for the identification and use of prognostic factors. *International Journal of Cancer* 1974; **13**:16–36.
36. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; **8**:467–475.
37. Senn SJ. Baseline comparisons in randomized clinical trials. *Statistics in Medicine* 1991; **10**:1157–1160.
38. Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. *Controlled Clinical Trials* 2000; **21**:330–342.
39. Armitage P. The analysis of data from clinical trials. *The Statistician* 1979; **28**:171–183.
40. Armitage P, Berry G, Matthews JNS. *Statistical Methods in Medical Research* 4th edn. Blackwell: Oxford, 2002.
41. Hernández AV, Steyerberg EW, Habbema JD. 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.
42. Negassa A, Hanley JA. The effect of omitted covariates on confidence interval and study power in binary outcome analysis: a simulation study. *Contemporary Clinical Trials* 2007; **28**:242–248.
43. Schoenfeld DA, Borenstein M. Calculating the power or sample size for the logistic and proportional hazards models. *Journal of Statistical Computation and Simulation* 2005; **75**:771–785.
44. Azzalini A. *The R* `sn` *package : the skew-normal and skew-t distributions (version 1.0-0).* Università di Padova, Italia, 2014. Available from: http://azzalini.stat.unipd.it/SN [Accessed on 6 February 2014].